FOLDING OF A MULTIDOMAIN PROTEIN, WITH THE HELP OF THE RIBOSOME AND A MOLECULAR CHAPERONE

By

Kaixian Liu

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

June 2019

Abstract

Multidomain proteins, containing several structural units within a single polypeptide, constitute a large fraction of all proteomes. Cotranslational folding is assumed to simplify the conformational search problem for large proteins, but the events leading to correctly folded structures remain poorly characterized. Additionally, how the ribosome and molecular chaperones promote efficient folding remains obscure. In this study, folding events of nascent elongation factor G, a five-domain protein that requires chaperones for folding *in vivo*, have been dissected by single-molecule optical tweezers. We found that the N-terminal G-domain of EF-G constitutes an independent folding unit. Upon *in vitro* refolding, it adopts two similar states that correspond to the natively and non-natively folded structures. The ribosome destabilizes both of them, suggesting terminal misfolding into highly stable, non-native structures is avoided. The first N-terminal two domains G and II will form misfolded off-pathway states in isolation, however, the ribosome and the chaperone trigger factor reduce inter-domain misfolding, permitting folding of the N-terminal G-domain. Successful completion of this step is crucial for folding of domain II. Unexpectedly, cotranslational folding does not proceed unidirectionally: emerging unfolded polypeptide denatures an already folded domain. Trigger factor, but not the ribosome, protects against denaturation. The chaperone thus helps multidomain proteins overcome inherent challenges during cotranslational folding. While the N-terminal two domains must fold in the order of their synthesis, the C-terminal domains do not fold sequentially. Domain III stability is dictated by the folded C-terminal *IV-V*, indicating a post-translational folding mechanism. In summary, as a paradigm for multidomain protein folding, this work discovered that domain-wise folding of nascent proteins can be reversed by denaturation

interactions with emerging polypeptide, which can be blocked by trigger factor. Trigger factor also cooperates with the ribosome to reduce misfolding. The chaperone therefore has a dual function in promoting efficient folding of multidomain proteins. Furthermore, the folding dependency of domain II and III on their neighboring domains supports the hypothesis that domains incapable of independent folding are stabilized by favorable interactions with neighboring domains, which could be a fundamental rule for multidomain protein folding.

Thesis advisor/Primary reader: Dr. Christian Kaiser

Secondary reader: Dr. Sarah Woodson

Acknowledgements

At this point, I am very grateful to write down this dissertation. There are so many people that have helped me along this journey from many different aspects. The following is an attempt to acknowledge my mentors, colleagues, friends, and family.

I first would like to thank my thesis advisor, Dr. Christian Kaiser. Christian is the best mentor I could have ever met. I learned how to be a good scientist from him, including how to design and perform experiments, analyze the data, present the data as well as think critically. In addition, I also learned how to achieve a healthy work-life balance from him. His scientific attitudes, communicational skills, enthusiasm and optimism have far-reaching influence on my scientific career.

I also would like to thank Dr. Vincent Hilser, who generously offered me a volunteer opportunity in his lab before I joined the biology department. Additionally, I want to thank Dr. Jing Li from the Hilser lab, who not only taught me the experimental skills, but also helped me tremendously to adapt to a new environment. I want to acknowledge Dr. Doug Barrick, Dr. Young-Sam Lee as well, who accepted me as a rotation student and provided terrific training opportunity for me. I would like to take the opportunity to express my particular thanks to my thesis committee members: Dr. Sarah Woodson, Dr. Doug Barrick, Dr. Taekjip Ha, and Dr. Greg Bowman. Their fantastic suggestions and kind help to my thesis project and career development are invaluable to me. I would also like to give special thanks to Dr. Xin Chen and Dr. Haiqing Zhao, who are great listeners and have shared a lot of great scientific insights as well as career guidance with me over the past few years.

I feel fortunate to have had a number of great friends and colleagues around. Riti Gupta, Anthony Milin, Dr. Sunyia Hussain, Kevin Maciuba, Xiuqi Chen, Olivia Osborn, Dr. Nagaraju Chada, and Nanda Rajasekaran, are all great lab mates. Joseph Rehfus and Elliot Mattson, it was great to have them as rotation students to help with my research.

Last but not the least, I want to thank all of my family members for them always being very supportive. Lijuan, I definitely would not come this far without you. My parents and grandparents always believe in me even more than I do. I hope my grandfathers who have left from us forever could have shared the joy that I am having now. Of course, I want to thank Raymond, my son, who is and will always be the reason why I keep improving myself.

Table of contents

List of tables

List of figures

Chapter I. Introduction

I-1: Protein folding

Most of the proteins in the cell have defined three-dimensional native structures. The biological roles of them are determined by their native structures. Biophysical studies have demonstrated that the amino acid sequence of a protein determines its structure and also encodes the pathway along which a protein folds from a linear polymer into a precisely ordered structure (Anfinsen, 1973). Over 50 years of studies have resulted in the view of the protein folding process as the diffusion over a free energy landscape, which is funnel shaped (Onuchic et al., 1997) with a narrow bottom (figure I-1). This conceptually simple theory has greatly helped to conceptualize the intricate folding process, and shaped our way of interpreting experimental data.

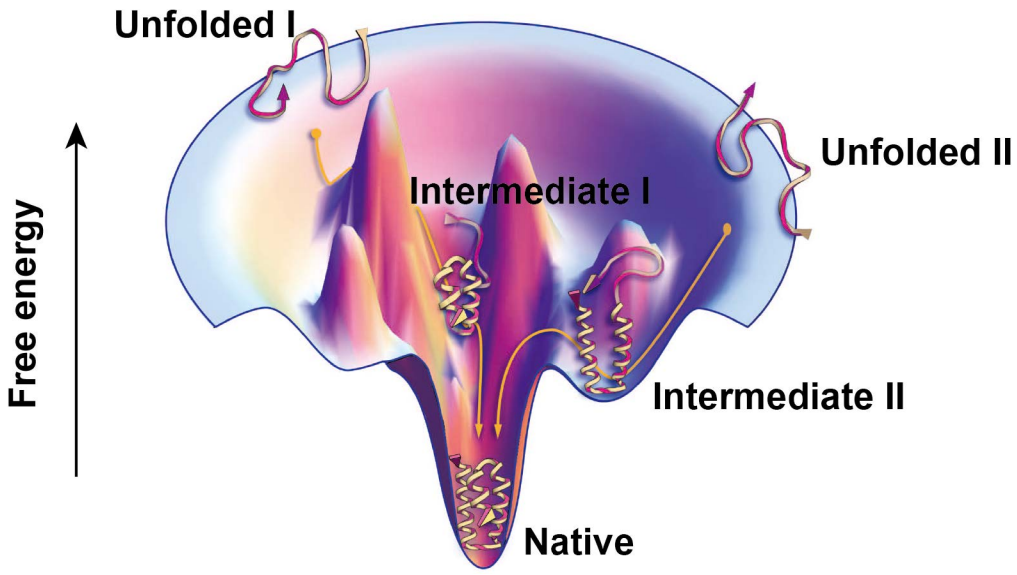


Figure I-1 Protein folding energy landscape theory. The protein starts from unfolded state, which has a very large conformational space. The native state occupies the free energy landscape minimum. There are high barriers separating the unfolded state and native state or intermediate states. As the overall free energy decreases, the conformational space gets narrower. Figure is adapted from Dill and MacCallum, 2012.

Despite this progress, there are still many open questions regarding to the molecular mechanism of folding, and it is not clear whether a consensus has been reached even for small single-domain protein folding (Sosnick and Barrick, 2011). For instances, the importance of residual structures in denatured states, the nature of folding steps and kinetic barriers as well as the extent of pathway diversity is still elusive (Sosnick and Barrick, 2011). Many of the physical parameters of protein structure are well understood: Hydrogen bonds are the main forces stabilizing secondary structures; van der Walls interactions are important for the tight packing of folded proteins; backbone angles have preferred distributions in different structures, as visualized in Ramachandran plots; electrostatic interactions are important for stabilizing the local and global structures; hydrophobic interactions are the main driving forces for proteins to form a core structure; chain entropy undergoes a large decrease when proteins fold, therefore modulating the overall stabilities of proteins (Dill and MacCallum, 2012). However, we are still unable to resolve Levinthal's paradox: How do proteins fold into precisely defined native structures quickly, given the vast number of possible conformations even for a small single-domain protein (Levinthal, 1969)? Furthermore, although several examples impressively illustrate our ability to design folded proteins *de novo* with atomic resolution (Dou et al., 2018; Kuhlman et al., 2003; Lu et al., 2018), it is still generally impossible to accurately predict a protein structure from its sequence. In the future, both the experimental methods and computational analysis need to be advanced to address these questions.

I-2: Multidomain protein folding

Protein folding has mainly been studied using small, single-domain proteins. The folding pathways of larger multidomain proteins, which represent a large fraction of proteomes

(Braselmann et al., 2013), remain poorly understood, partly because side reactions such as aggregation complicate folding measurements. Unlike single-domain protein folding, frustration from strong inter-domain interaction can exacerbate in multidomain proteins (Zheng et al., 2013). Borgia et al., using single-molecule fluorescence to study folding of tandem I27 domains, found that misfolding happens in a sequence dependent manner, which is that the higher sequence similarity between tandem domains, the higher likelihood of misfolding, suggesting an evolutionary strategy to minimize interdomain misfolding (Borgia et al., 2011). Further detailed studies have shown that the lifetime of misfolded states seem to correlate with the sequence identity between neighboring domains, with longer lifetime for higher sequence identity (Borgia et al., 2015). Interestingly, ensemble measurement with spectrin domains R15R16 has shown that a neighboring domain will not affect the folding pathway of the other domain (Batey and Clarke, 2008). Computational genomic studies have suggested that the folding of domains becomes independent when the interface is small and loosely packed (Han et al., 2007). This finding is consistent with another computational study showing that the binding affinity between domains are positively correlated with buried surface areas of the interface (Chen et al., 2013). Recently, folding of a complex multi-domain protein Hsp90 has been investigated by single-molecule optical tweezers (Jahn et al., 2016). This study revealed that a small stretching force applied to the polypeptide chain can accelerate folding, presumably by reducing non-productive long-range contacts between non-native domains. The authors suggest that a similar effect might be achieved by cotranslational folding, since cotranslational folding enables domain-wise folding of multidomain proteins, therefore preventing inter-domain misfolding. This study highlights the importance of the ribosomal environment for multidomain protein folding. Indeed, it has been

demonstrated that interactions with the cellular biosynthetic machinery to coordinate protein synthesis and folding are particularly important to optimize the folding efficiency of multidomain proteins (Kim et al., 2015).

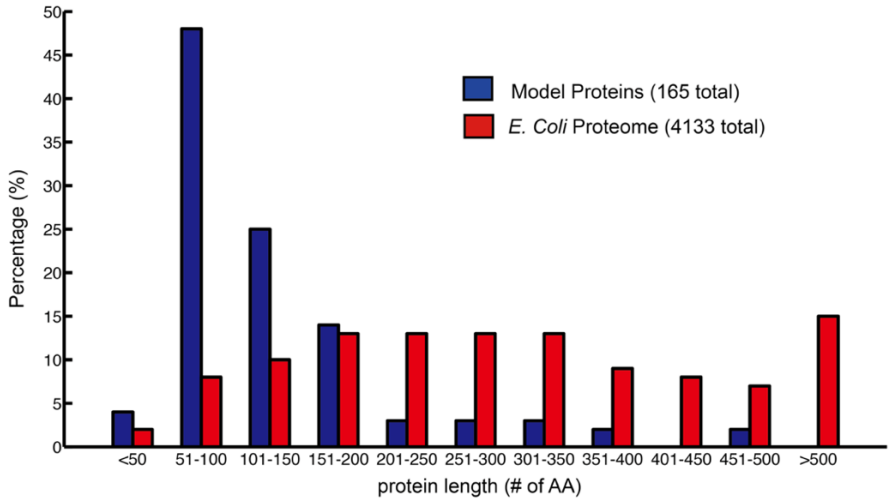


Figure I-2 Multidomain proteins are understudied. Protein size distribution of *Escherichia coli* proteome (red) and model proteins (blue) whose folding has been characterized. Figure is adapted from (Braselmann et al., 2013).

I-3: Molecular chaperones

It is well established that many proteins require the assistance of molecular chaperones for efficient folding in the cell (Hartl et al., 2011). Specialized chaperone systems interact with ribosome-bound nascent chains as they emerge from the ribosomal exit tunnel (Kristensen and Gajhede, 2003). The bacterial chaperone TF is targeted to the polypeptide exit site in the large ribosomal subunit where it interacts with nascent polypeptides before their synthesis is complete. TF is the first molecular chaperone that nascent polypeptide chain encounters in bacteria. TF consists of an N-terminal ribosome binding domain, a central PPIase domain, and a C-terminal domain that was suggested to comprise the main substrate binding site (Ferbitz et al., 2004). It was found that the association of TF with nascent polypeptides is regulated by ribosome binding (Kaiser et al., 2006), providing a mechanism for nucleotide-independent regulation. Ribosome

profiling experiments have suggested that TF binds to the nascent chain only after more than 100 amino acids have been synthesized, and that engagement of the nascent chain with multiple TF molecules is unlikely (Oh et al., 2011). However, these results are not consistent with results from the FRET studies on translating ribosomes (Kaiser et al., 2006) and NMR measurements (Saio et al., 2014). It remains unclear how trigger factor itself tunes its PPIase and holdase activities to optimize the folding yield of nascent polypeptides.

While trigger factor is unique to bacteria, the nascent chain associated complex (NAC) is thought to act as functionally equivalent chaperone for nascent polypeptides in eukaryotes (Preissler and Deuerling, 2012). The universally conserved Hsp70 system also associates with nascent polypeptides. It has been proposed to engage the nascent chain at a later stage (Hartl and Hayer-Hartl, 2009) and thus act later during folding. However, the bacterial Hsp70 system can functionally replace TF (Genevaux et al., 2004) and it is not clear how the functions of these chaperones are coordinated. Hsp70 is regulated in an ATP-dependent fashion by a co-chaperone and nucleotide exchange factors and represents one of the best-understood chaperones (Hartl and Hayer-Hartl, 2009). How nascent chain-binding chaperones change the folding energy landscape of their client proteins is still poorly understood. Previous studies indicate that TF and DnaK have partially overlapping functions, but it is still elusive to what degree their functions are distinct and how they work together (Genevaux et al., 2004). Similarly, how contributions from chaperones and the ribosome are coordinated is incompletely understood (Hoffmann et al., 2010).

I-4: Elongation factor G (EF-G)

EF-G (eEF-2 in eukaryotes) is an abundant, universally conserved GTPase that facilitates mRNA and tRNA translocation. EF-G is composed of five domains, including an N-terminal GTPase domain termed G-domain, followed by domains II, III, IV and V. Domain IV is discontinuous in the primary structure, and hence subdomains IVa and IVb flank domain V (Figure I-3A). Crystal structures (Figure I-3B) of EF-G (AEvarsson et al., 1994; Czworkowski et al., 1994) have revealed that the G-domain is structurally similar to but larger than other small GTPases, such as p21 or ras, due to the insertion of a G' domain in EF-G. Interestingly, a previous study found that an artificial multidomain protein, fusing ras to DHFR, was able to fold only through a cotranslational mechanism, but would result in interdomain misfolding if folding happened post-translationally (Netzer and Hartl, 1997). The modular architecture and structural complexity, which is missing in previous studied tandem repeat multidomain proteins (Batey et al., 2006; Borgia et al., 2011; Law et al., 2003), make EF-G an attractive model for studying multidomain protein folding. Efficient folding of EF-G in the cell depends on molecular chaperones: The protein aggregates, indicating severe folding defects, upon deletion of TF and the Hsp70 system (Deuerling et al., 2003; Vorderwulbecke et al., 2004), the two major nascent chain-binding proteins in *E. coli*. Proteomics studies with *E. coli* also found that EF-G is a natural client of DnaK, which is a Hsp70 family member protein (Calloni et al., 2012). Based on these findings, EF-G is a promising model protein for studying the contributions of vectorial synthesis and of chaperone interactions to multidomain protein folding.

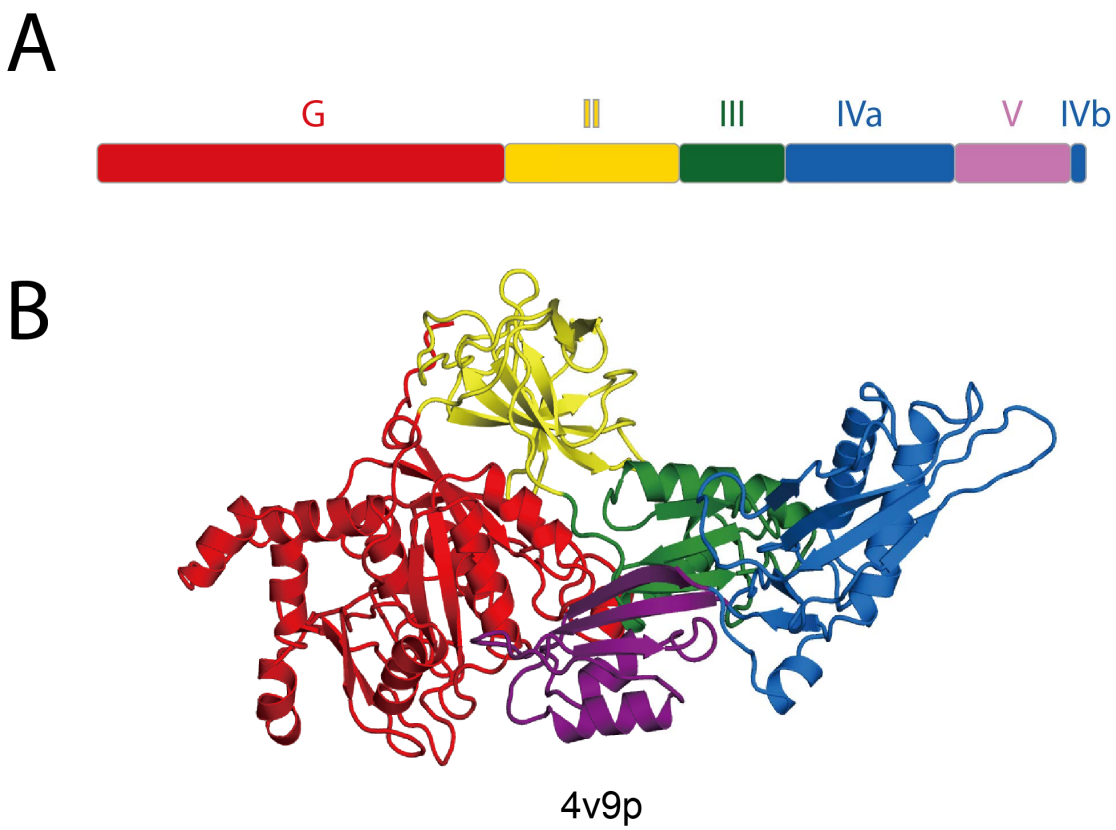


Figure I-3 Primary structure diagram and 3D structure of EF-G.

I-5: Aim of this study

The aims of the present study are to determine (1) how the folding of individual domains is coupled in a natural multidomain protein EF-G, (2) how the ribosome contributes to the folding of EF-G and (3) how the molecular chaperone trigger factor (TF) changes the folding free energy landscape of EF-G.

Using single-molecule force spectroscopy to dissect folding and unfolding of full-length EF-G, its truncations, and individual domains, we found that the N-terminal two domains must fold sequentially, suggesting a domain-wise cotranslational folding mechanism. However, domain III folding may be post-translational, due to that the domain III stability highly depends on C-terminal domains IV and V. Inter-domain misfolding seems to be a general mechanism that

slows down EF-G folding. An unexpected unfolding pathway was discovered in the N-terminal two domains, which is that the folded G-domain can be denatured by the unfolded domain II.

By investigating the folding of nascent EF-G chains of defined lengths in the context of the ribosome, we found that how the ribosome regulates protein folding is highly stage dependent, meaning that the ribosome can both decelerate and accelerate distal domain folding depending on how far away the domain is from the ribosome. However, the underlying mechanism of accelerating distal domain folding could be the same as decelerating, with the difference being that acceleration is achieved by interactions between the ribosome and the approximate domain which can prevent inter-domain misfolding.

Using force spectroscopy to study the folding of EF-G nascent chains in the presence and absence of the ribosome-binding chaperone TF, we revealed that TF cooperates with the ribosome to reduce inter-domain misfolding, thus promoting folding of the distal N-terminal domain. The denaturation can be suppressed by TF, but not by the ribosome. Therefore, we have a better understanding how nascent chain binding chaperone TF reshapes folding of EF-G during translation.

Chapter II: Review of recent progress in understanding cotranslational protein folding

By

Christian Kaiser and Kaixian Liu

This work has been published as:

Kaiser, C.M., and Liu, K. (2018). Folding up and Moving on—Nascent Protein Folding on the Ribosome. *J. Mol. Biol. 430*, 4580–4591.

Chapter II. Review of recent progress in understanding cotranslational protein folding

II-1: Cotranslational protein folding

Proteins are linear polymers synthesized by the ribosome that generally have to fold into specific three-dimensional structures to become biologically active. Recently, it has been realized that the ribosome itself modulates early folding events (Thommen et al., 2017), helping to set the nascent polypeptide on the correct path for successful folding. The ribosome is the central player in cellular protein synthesis, polymerizing amino acids in the order dictated by the genetic information in the messenger RNA (mRNA). For every codon in the mRNA, the ribosome helps to select the appropriate amino-acylated transfer RNAs (tRNAs) and subsequently catalyzes the addition of the selected amino acid to the nascent protein. By translating the information in the mRNA one codon at a time, the ribosome synthesizes polypeptides in a directed manner, adding amino acids to the carboxyl-terminus (C-terminus) of the nascent polypeptide. As a consequence, proteins emerge in a vectorial fashion, starting at the amino-terminus (N-terminus), as the ribosome moves along the mRNA (Figure II-1).

Protein synthesis by the ribosome is highly processive. Full-length translation products are released only when a stop codon in the mRNA is reached. While stalling of ribosomes during elongation can lead to ribosome rescue (Himeno et al., 2014) or dissociation of the incomplete peptide (Heurgué-Hamard et al., 1998), the growing nascent polypeptide remains stably anchored to either the A-site or the P-site tRNA during normal elongation. At the same time, polypeptide elongation is relatively slow, proceeding at a rate of up to approximately 20 amino acids per second (Young and Bremer, 1976) in bacteria and even slower in eukaryotes as

estimated to be about 6 amino acids per second by both pulse-chase studies (Boström et al., 1986) and ribosome profiling data (Dao Duc and Song, 2018; Ingolia et al., 2011). Synthesis of even a small domain of 100 amino acids thus requires several seconds.

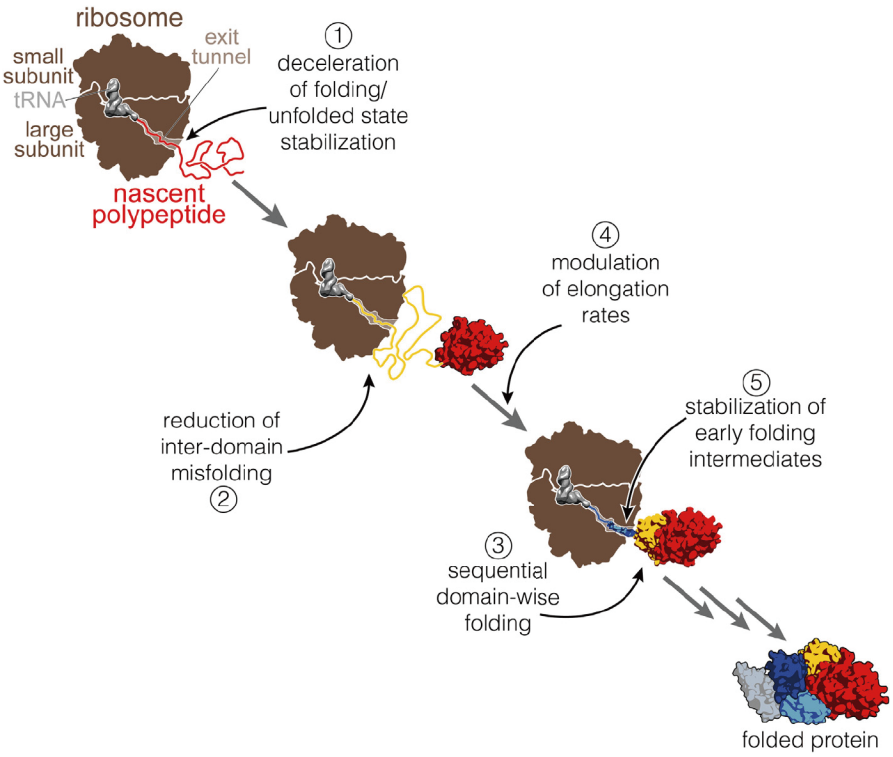


Figure II-1 Modulation of nascent chain folding by the ribosome. The cartoon depicts ways in which the ribosome and the process of translation can affect the folding of nascent polypeptides. (1) Interactions with the ribosome surface reduce folding rates and stabilize the unfolded state. (2) Presumably similar interactions might help to reduce misfolding among emerging domains in multi-domain proteins. (3) Sequential domain-wise folding reduces the complexity of the conformational search problem. (4) Local variations in elongation rate tune synthesis of folding, promoting the formation of native structures. (5) Structures that are unstable by themselves can be stabilized in specific regions of the ribosome exit tunnel that spans the large subunit.

Folding, on the other hand, can be much faster. While large proteins may take several minutes to fully complete folding, the formation of intermediates or the folding of domains is generally much faster, with time constants often in the millisecond range (Kubelka et al., 2004). Therefore, folding can in principle begin before synthesis is complete, as has been noted more than half a century ago (Phillips, 1967). Cotranslational folding of ribosome-bound nascent proteins could indeed be observed experimentally by assaying the catalytic activity (Nicola et al., 1999), the acquisition of protease-resistant structures (Frydman et al., 1999), or the binding of conformation-specific antibodies (Clark and King, 2001). These experiments also suggested that the particular environment near the ribosome and the vectorial process of polypeptide synthesis

affect what structures the nascent polypeptide populates (Clark and King, 2001; Frydman et al., 1999). Moreover, due to a variety of factors, the elongation rates along a given mRNA are not uniform (Rodnina, 2016): some codons are translated faster than others, resulting local minima and maxima at specific positions during polypeptide synthesis. The modulation of elongation rates has been shown to affect the biological function (Kimchi-Sarfaty et al., 2007; Zhou et al., 2013) or biochemical properties (Sander et al., 2014; Zhang et al., 2009) of the translation product. More recently, it has become possible to directly relate changes in elongation rate to the folding and structure of the newly synthesized polypeptide (Buhr et al., 2016; Kim et al., 2015).

Before being exposed to the cellular environment, nascent proteins pass through a long, narrow tunnel that spans the large subunit of the ribosome and can accommodate between 30 and 40 amino acids of a polypeptide in a largely unstructured conformation. While limited folding can occur within the interior of the ribosome, (Lu and Deutsch, 2005; Nilsson et al., 2015; Woolhead et al., 2004), the narrow exit tunnel does not allow extensive tertiary structure formation. Regulatory nascent peptides can adopt specific conformations within the exit tunnel and, through interactions with rRNA or protein components of the tunnel, regulate ribosome function (Gupta et al., 2016; Ishii et al., 2015; Ito and Chiba, 2013; Su et al., 2017; Wilson et al., 2016). The ribosome-nascent chain interactions are required to stabilize these conformations (Lucent et al., 2010; Woolhead et al., 2006), which are hence not stable outside the tunnel. Once the nascent polypeptide emerges from the tunnel, folding is less restricted, although close proximity of the ribosome limits the space available to the polypeptide and thus, in principle, the accessible conformations (Mittal and Best, 2008).

Experimentally measuring ribosome-nascent chain interactions and their effects on folding is challenging. Mechanistic studies of protein folding traditionally utilize thermal or chemical denaturation of the protein of interest and simultaneously monitoring folding and unfolding using optical spectroscopy. The biochemical complexity of the ribosome, which in bacteria is composed of three ribosomal RNA (rRNA) molecules and more than 50 proteins, precludes the utilization of intrinsic spectroscopic probes that are commonly employed to follow folding and unfolding (e.g., tryptophan fluorescence). Moreover, the traditional approach of adding and removing denaturants to study folding cannot be used, because globally acting denaturant destabilizes not only nascent polypeptide structure, but also the ribosome itself. Over the last years, novel approaches have been developed to overcome these limitations. Most experiments to date have utilized ribosome-nascent chain complexes (RNCs) in which translation is stopped at precisely defined nascent chain lengths, either through the use of a genetically engineered arrest peptide from the SecM protein (Nakatogawa and Ito, 2002) or by using stop codon-less (nonstop) mRNA templates. Recently, a carefully designed *in vitro* translation system (Mittelstaet et al., 2013) has opened exciting avenues for simultaneously measuring of polypeptide elongation and folding.

II-2: Biochemical approaches

Several biochemical approaches have been used to map the onset of folding as a function of nascent chain length and the effect of the ribosome on folding and stability (Figure II-1). Hoffmann et al. employed an experimental strategy relying on structure-dependent disulfide bond formation between strategically placed cysteine residues. These experiments revealed distance-dependent destabilization of nascent chain structure by the ribosome that was similar for

a set of model proteins (including the Scr homology 3 domain of α-spectrin, barnase, and β-lactamase) (Hoffmann et al., 2012). Interestingly, the effect is amplified by the ribosome-binding chaperone trigger factor (Hoffmann et al., 2012), suggesting one mechanism by which the chaperone may cooperate with the ribosome to modulate nascent protein folding.

To obtain quantitative information on the destabilization caused by the ribosome, Samelson et al. (Samelson et al., 2016) employed "pulse proteolysis" (Park and Marqusee, 2005), a refined version of the limited proteolysis approach that had been exploited in early experiments studying the folding of nascent firefly luciferase (Frydman et al., 1999). Instead of incubating RNCs with low concentrations of protease, samples are "pulsed" with a high protease concentration for a brief period of time. The pulse duration is short relative to the mean lifetime of the folded state of the protein of interest. Consequently, only the fraction that populates the unfolded state during the pulse is digested, whereas the folded state population is protected. The equilibrium fraction of folded protein determined in this way directly yields its global stability. For the three globular proteins analyzed in this study (versions of dihydrofolate reductase, ribonuclease H and barnase), the ribosome caused a significant destabilization when the folded units were in close proximity to the ribosome. Lengthening the C-terminal extension from 35 to 55 amino acids restored stability, confirming earlier observations that the effect of the ribosome is restricted to its immediate vicinity (Cabrita et al., 2016; Holtkamp et al., 2015; Kaiser et al., 2011; Nilsson et al., 2015). Destabilization of folded structures in the immediate vicinity of the ribosome, presumably by interactions stabilizing the unfolded state, therefore emerges as a general phenomenon.

The SecM arrest peptide is frequently used as a tool to prepare stably stalled RNCs in vitro (Schaffitzel and Ban, 2007) or in vivo (Cabrita et al., 2009). The mechanism underlying SecM-mediated arrest has also made this sequence a useful tool for studies of nascent chain folding dynamics. Biochemical and genetic studies have suggested that mechanical force, generated by the Sec translocon (Butkus et al., 2003; Nakatogawa and Ito, 2002) or insertion of transmembrane helices into lipid bilayers (Ismail et al., 2012), accelerates release of SecM-induced elongation arrest. Single-molecule force-spectroscopy experiments established that force is indeed sufficient to release arrest (Goldman et al., 2015). Folding of the nascent chain into a stable structure near the ribosome is thought to generate a pulling force that accelerates arrest release (Goldman et al., 2015; Nilsson et al., 2016). When arrest release is coupled to an experimental readout, this phenomenon can be exploited to define the nascent chains lengths at which folding occurs (Figure II-2). Systematically varying the distance between a folding domain and the ribosome confirmed that folding into stable structures occurs at a distance of approximately 30 to 40 amino acids from the peptidyl transferase center. Folding-mediated release of SecM arrest has been demonstrated in live *E. coli* cells (Goldman et al., 2015) (Figure II-2). It should be noted that the exact mechanism of force generation by nascent chain folding is not completely understood, and that arrest release is slow relative to normal elongation rates. SecM arrest release experiments are therefore difficult to interpret with certainty in terms of folding kinetics, but it seems clear that this experimental approach robustly detects the cotranslational formation of stable structures in the nascent polypeptide.

Several studies have employed folding-mediated release of SecM arrest to detect folding of nascent membrane (Cymer and von Heijne, 2013; Ismail et al., 2012) and soluble proteins

(Farías-Rico et al., 2017; Goldman et al., 2015; Marino et al., 2016; Nilsson et al., 2015, 2016, 2017). Nilsson et al. utilized the approach, in combination with cryo-electron microscopy, to map the onset of co-translational folding (Nilsson et al., 2017). Different versions of the three-helix bundle spectrin domain begin to fold right around the position where they fully emerge from the ribosome exit tunnel. Analyzing several spectrin variants revealed that, surprisingly, the onset of nascent chain folding does not correlate with their intrinsic stabilities, folding rates or folding pathways of the isolated polypeptides. Instead, the results suggest that interactions with the ribosome can alter the folding pathway of a nascent chain compared to the same polypeptide in isolation. As a consequence, the folding pathways determined with isolated proteins may not always reflect the sequence of events during cotranslational folding, and interactions with the ribosome can change the folding properties in ways that are at present not predictable. While the arrest peptide assays cannot resolve folding intermediates at the resolution provided by biophysical measurements, they nevertheless represent a very useful approach for defining "waypoints" for the cotranslational folding of nascent proteins in vitro and in living cells.

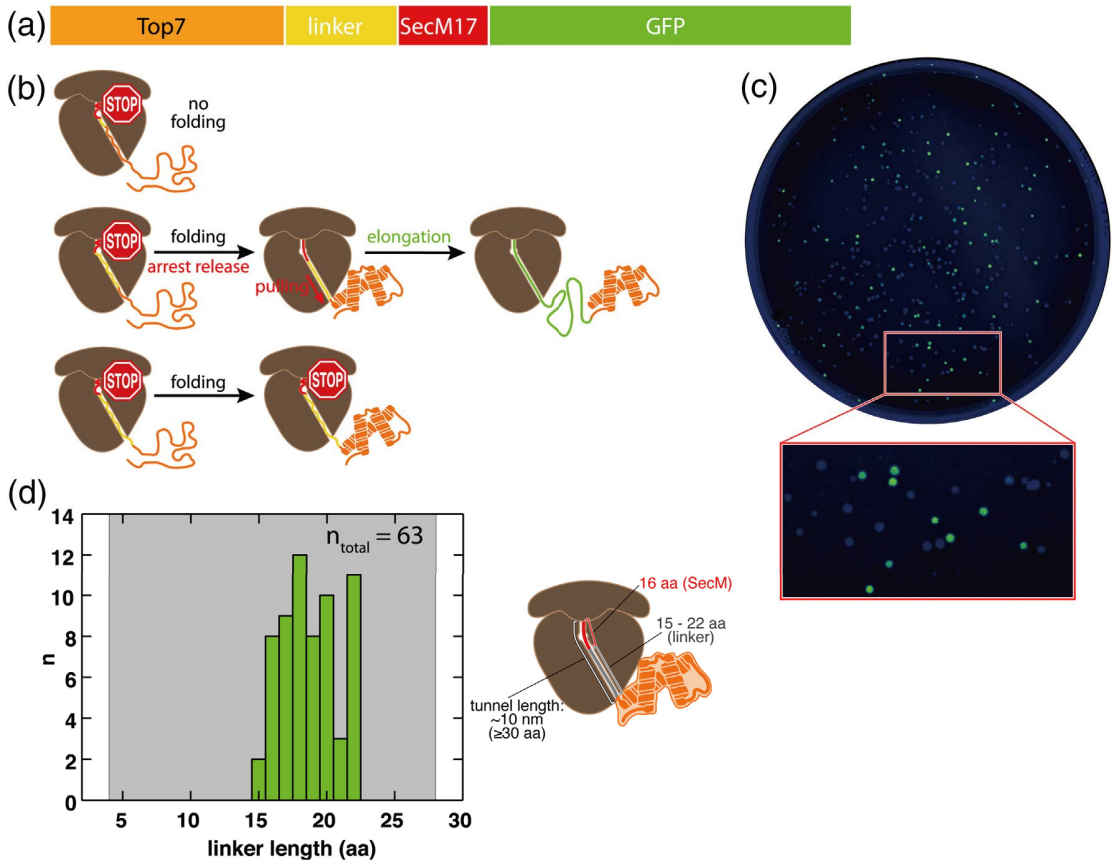


Figure II-2 The SecM arrest peptide as a tool to detect nascent chain folding. (A) Bar diagram of the construct used to assess nascent chain folding. The SecM arrest peptide (red) is fused to an independently folding domain (Top7), orange) by linkers of variable lengths (yellow). The downstream green fluorescent protein (GFP, green) is translated only after SecM arrest has been released. Varying the linker length from 4 aa to 28 aa creates a linker "library". (B) Schematic illustrating the expected translation outcome of very short linkers (<<15 aa, top), intermediate length linkers (~15 aa, middle) and long linkers (>> 15 aa, bottom). Force is generated by nascent chain folding at intermediate linker lengths, resulting in elongation restart. (C) Agar plate with colonies transformed with the linker library under UV illumination. A fraction of the colonies exhibits green fluorescence, the remaining colonies exhibit weak blue autofluorescence. (D) Distribution of linker lengths obtained by sequencing plasmid DNA from fluorescent colonies (green bars). The recovered length range represents a well-defined subset of all linker lengths present in the library (grey shaded area). The cartoon on the right illustrates that the lengths supporting elongation restart (31 to 38 amino acids, combining SecM and the flexible linker) are within the range expected based on the tunnel length. (modified from Goldman et al. 2015).

II-3: Fluorescence spectroscopy

Incorporation of exogenous fluorophores into nascent polypeptides (Crowley et al., 1994) has proven to be a useful tool for direct observations of nascent chain dynamics and folding through a variety of approaches. Measurements of the time-resolved anisotropy decay of a

bodipy fluorophore at the N-terminus of ribosome-tethered apo-myoglobin showed reduced conformational flexibility, as compared to the free polypeptide after release from the ribosome (Ellis et al., 2008, 2009). Experiments with a disordered protein further suggested that electrostatic interactions with the ribosome surface increase nascent chain dynamics (Knight et al., 2013), which may explain the destabilization of folded structures in ribosome-bound nascent chains that has been observed with other approaches (Cabrita et al., 2016; Kaiser et al., 2011; Samelson et al., 2016).

By introducing two fluorophores into the nascent polypeptide, folding can be observed through fluorescence resonance energy transfer (FRET). Khushoo *et al.* (Khushoo et al., 2011) combined a genetically engineered N-terminal cyan fluorescent protein (CFP), serving as the FRET donor, with a small organic fluorophore incorporated at various positions in the nascent chain, serving as the FRET acceptor, to map the folding of the first nucleotide binding domain (NBD1) in nascent cystic fibrosis transmembrane conductance regulator (CFTR). FRET measurements indicated that NBD1 folded into a native-like structure even before the entire domain had emerged, suggesting the population of a folding intermediate comprised of the N-terminal subdomain. Notably, binding of ATP stabilizes the intermediate, facilitating subsequent folding as more of the polypeptide becomes available during synthesis. Ligand binding therefore promotes the formation of a cotranslational folding intermediate, driving vectorial folding during synthesis.

Since many proteins have binding sites for small molecules, cotranslational stabilization of partially folded structures by ligand binding may be a general mechanism for facilitating cotranslational folding. For a system with suitable properties, specific ligand binding itself can

report on nascent chain folding and its modulation by the ribosome. Experiments exploiting the quenching of flavin mononucleotide upon binding to flavodoxin have recapitulated the previously observed destabilization of folded structures and indicated that stabilizing interactions due to ligand binding can shift the conformational equilibrium (Houwman et al., 2016). Specific binding interactions that drive nascent protein folding are not limited to small molecule ligands. Hetero-oligomeric complex formation can similarly begin while one subunit is still being synthesized, resulting in complex assembly with increased efficiency in vitro (Fedorov and Baldwin, 1995) and in vivo (Shieh et al., 2015).

The careful design of an in vitro translation system that combined robust elongation rates with efficient incorporation of exogenous spectroscopic probes has recently enabled real-time measurements of cotranslational folding (Holtkamp et al., 2015). FRET measurement, using donor and acceptor dyes incorporated into the nascent polypeptide (Figure II-3), indicated the formation of a compact intermediate early during translation that is not significantly populated in the isolated polypeptide. Based on nascent chain length, the intermediate forms within the distal part of the ribosome exit tunnel. Limited proteolysis and photo-induced electron transfer (PET) measurements confirmed the co-translational population of the intermediate that converts into the native structure upon chain elongation or release from the ribosome. More recent work (Mercier and Rodnina, 2018) based on analyzing intrinsic fluorescence intensities of different fluorophores from FRET- and PET- dependent time courses suggests that the N-terminal α-helical domain of universally conserved N^5-glutamine methyltransferase HemK forms at least four intermediates through a “helix docking” mechanism, which is rate-limited only by translation. These elegant experiments demonstrate that the ribosomal environment has a

profound effect on nascent chain structure and pave the way to investigating the effects of elongation rate modulation on cotranslational folding pathways in mechanistic detail.

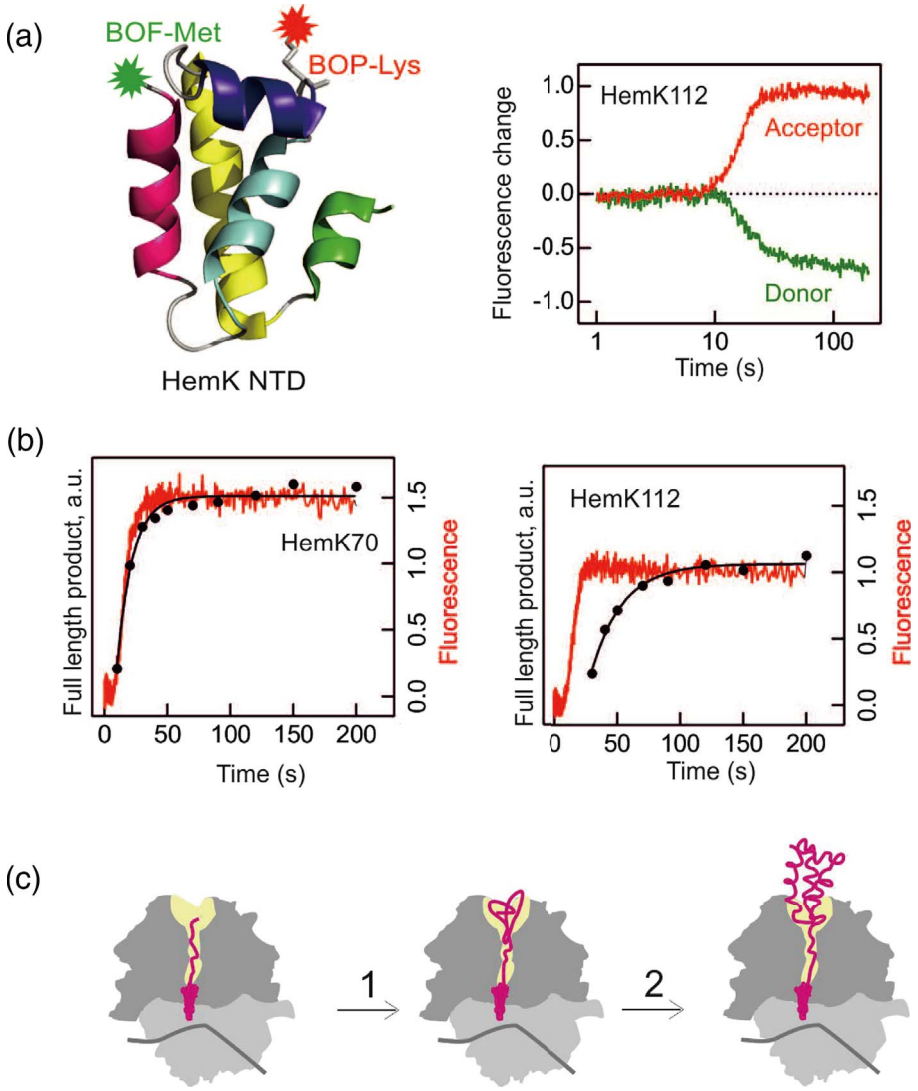


Figure II-3 Cotranslational folding monitored in real-time. (A) Crystal structure (left) of the HemK N-terminal domain (residues 1-73 of the full-length protein). Green and red stars indicate the positions (amino acids 1 and 34) of the donor (BOF-Met) and acceptor (BOP-Lys) dyes incorporated into the polypeptide. Changes in BOF-MeT (green) and BOP-Lys (red) fluorescence during HemK synthesis (right). The drop in donor and simultaneous increase in acceptor fluorescence at 10 seconds indicate that energy transfer due to folding of the nascent polypeptide. (B) Time course of an acceptor fluorescence and translation product accumulation. For translation templates of varying length, the acceptor fluorescence rapidly increases after 10 seconds. For a construct containing the N-terminal 70 amino acids of HemK (HemK70), fluorescence increase and product accumulation occur simultaneously. When the entire N-terminal domain emerges from the ribosome (HemK112), the fluorescence increase precedes the completion of the synthesis, demonstrating cotranslational folding. The fluorescence signal in HemK70 is higher than that of shorter or longer nascent chains, suggesting that the ribosome stabilizes a compact structure at this nascent chain length. (C) Cartoon of cotranslational HemK folding. The peptidyl-tRNA is shown in

magenta. A compact structure is stabilized in the distal portion of the ribosomal exit tunnel (yellow) that spans the large ribosomal subunit (dark grey) (step 1). After further elongation, the nascent polypeptide rearranges into a native-like structure. (from Holtkamp et al. 2015)

II-4: NMR spectroscopy

FRET measurements provide a sensitive readout of distance changes with good time resolution, but yield only one-dimensional information about the structure formed by the nascent polypeptide. Nuclear magnetic resonance (NMR) spectroscopy not only provides a sensitive measurement of protein structure, but also has the potential to resolve heterogeneous ensembles of states that are likely populated during folding. While not probing cotranslationally formed structures directly, NMR experiments with human gamma-B crystallin showed that synonymous changes in the coding sequence result in the formation of distinct structures of polypeptides with identical sequences, demonstrating the importance of local translation rates for correct folding (Buhr et al., 2016).

Selective isotope labeling of newly translated polypeptides (Buhr et al., 2016; Hsu et al., 2007) and a recently developed methyl TROSY based pulse sequence (Tugarinov and Kay, 2003) have enabled high-resolution measurements of nascent chain structure in the context of ribosome (Cabrita et al., 2016). Combining constraints from these measurements with molecular dynamics simulations yielded structural ensembles of the filamin domain 5 (FLN5) of the Dictyostelium discoideum gelation factor (Cabrita et al., 2016). Characterizing the structure of nascent chains arrested at several defined positions revealed that FLN5 attains its native structure only after at least 47 residues of the following domain (FLN6) have been translated, well after it has emerged from the exit tunnel. In the absence of the ribosome, isolated FLN5 folds even when lacking four C-terminal residues. Ribosome-nascent chain interactions therefore destabilize the native FLN5 structure, effectively inhibiting its folding until a substantial fraction of the

following FLN6 domain has been extruded. The NMR experiments also revealed contacts between the partially synthesized, disordered FLN6 domain and ribosomal RNA and proteins, in particular L23 and L24. These measurements are thus beginning to reveal, in molecular detail, the cotranslational interactions that modulate the folding landscape of nascent proteins. Consistent with previous studies (Kaiser et al., 2011; Knight et al., 2013), experiments with adding ribosome in trans to unfolded FLN5Δ12 and unfolded variant FLN5Y719E resulted in similar signal reduction as in corresponding RNCs, but not for full-length folding competent FLN5, further supports the conclusion that the ribosome selectively interact with unfolded nascent chain.

NMR spectroscopy of ribosome-bound nascent chains is technically very demanding, and the approach is still being developed to its full potential. It promises to be extremely powerful because it has the capability to resolve the (equilibrium) structures populated during early cotranslational folding at atomic resolution (Cabrita et al., 2016), which is not currently possible with other approaches. Simultaneously, such measurements yield information on their thermodynamic stability. Due to the long acquisition times, actively elongating ribosome-nascent chain complexes cannot be analyzed, limiting the approach to equilibrium measurements with stalled nascent chains. Complementing high-resolution structural information from NMR experiments with the temporal resolution attainable by fluorescence measurements or force spectroscopy (see below) has the potential to yield transformative insights into the dynamics of nascent protein folding.

II-5: Force Spectroscopy

Cotranslational folding is intrinsically a non-equilibrium process. Defining its kinetics is therefore crucial. Single-molecule approaches are very powerful tools for dissecting protein folding landscapes, yielding not only information about the structures that are populated along folding pathways, but also about kinetic rates (Schuler and Hofmann, 2013; Woodside and Block, 2014). They resolve the inherent stochasticity of folding and the transient, potentially heterogeneous states populated en route to the native state. Single-molecule force spectroscopy with optical tweezers has unique capabilities for characterizing cotranslationally formed states both kinetically and thermodynamically (Zoldak and Rief, 2013). Mechanical force acts as a denaturant, "tilting" the free energy landscape of a protein and favoring states that are more extended and less structured than the native state (Bustamante et al., 2004). In contrast to chemical denaturants that globally act on all molecules in solution, force is applied locally and can therefore be used to selective perturb the stability of nascent proteins while leaving the ribosome intact. Manipulation of single molecules also circumvents the problem of inter-molecular aggregation, which complicates experiments with unfolded or partially structured states that are inevitably populated during polypeptide elongation.

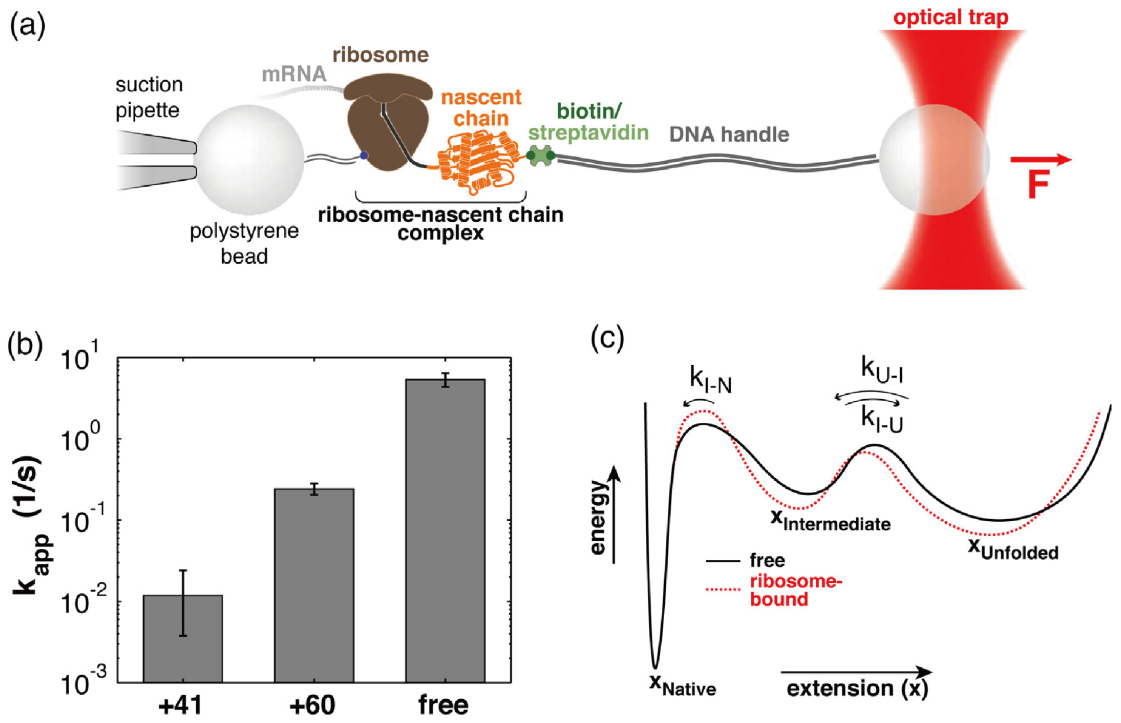


Figure II-4 Folding of T4 lysozyme on the ribosome. (A) Schematic of the molecular assembly for optical tweezers experiments. A ribosome-nascent chain complex is tethered between two polystyrene beads by means of DNA handles attached to the large ribosomal subunit and the N-terminus of the nascent polypeptide. Force is applied to the tethered complex by moving the optical trap relative to the pipette. The T4 lysozyme polypeptide (orange cartoon) is linked to the ribosome by means of a C-terminal Ser/Gly linker (grey). (B) In close proximity to the ribosome (41 amino acid C-terminal linker, "+41"), the apparent rate of T4 lysozyme folding is reduced more than 100-fold compared to the isolated protein in the absence of the ribosome ("free"). Extending the linker to 60 amino acids ("+60") partially restores the folding rate, indicating that the effect of the ribosome on nascent chain folding rates strongly depends on distance. (C) Schematic of the folding energy landscape obtained from optical tweezers experiments for free (black) and ribosome-bound (red) T4 lysozyme. The ribosome stabilizes both the unfolded and an intermediate state. The equilibrium between these two states is not affected by the ribosome, but the barrier for the final transition to the native state is increased, explaining the observed deceleration of folding. The ribosome therefore modulates a specific step in the folding pathway. (from Kaiser et al. 2011)

Optical tweezers have been utilized to study how the ribosome modulates the folding of T4 lysozyme (Kaiser et al., 2011) (Figure II-4), a small protein whose folding in isolation has been extensively studied (Mercier and Rodnina, 2018). When the T4 lysozyme polypeptide is kept in close proximity to the ribosome by means of a 41 amino acid long C-terminal linker, folding is slowed down more than 100-fold compared to the protein in the absence of the ribosome (Kaiser et al., 2011). Extending the linker to 60 amino acids partially restored folding

rates, suggesting that interactions with the ribosome modulate folding. The formation of an obligatory folding intermediate, most likely corresponding to the C-terminal subdomain of the protein, is not affected by the ribosome, suggesting that the ribosome can modulate specific steps along the folding pathway, rather than globally destabilizing any folded structure. Whereas C-terminally truncated T4 lysozyme adopts bona fide misfolded states and aggregates in isolation, both misfolding and aggregation are suppressed in ribosome-bound polypeptides (Kaiser et al., 2011). A potential function of the ribosome may thus be to prevent premature folding of an emerging domain until sufficient polypeptide has been synthesized to allow productive folding.

Tuning stabilities and folding rates of nascent polypeptides may be particularly important for larger, multidomain proteins. Technical challenges have largely precluded a mechanistic understanding of how these complex proteins navigate their intricate folding landscapes. The formation of misfolded, off-pathway species that has been directly observed in single-molecule experiments (Jahn et al., 2016; Liu et al., 2017) complicates folding to the native structure. Domain-wise folding during synthesis is a straightforward mechanism of reducing the complexity of the conformational search problem for multi-domain proteins. Cotranslational folding is therefore likely to be a key aspect in their biogenesis. Liu et al. have recently reported a first step toward dissecting the cotranslational folding pathway of the five-domain protein elongation factor G (EF-G) (Liu et al., 2017, also see chapter IV). The N-terminal G-domain of EF-G folds into a stable structure both in isolation and on the ribosome. As was observed with T4 lysozyme (Kaiser et al., 2011), G-domain folding proceeds more slowly on the ribosome (Liu et al., 2017). The kinetic modulation of nascent chain folding by the ribosome, presumably by sequestering part of the polypeptide into interactions with the ribosome surface, might thus be a

general aspect of cotranslational folding. In the context of a multidomain protein, such interactions could also serve to limit the entanglement of several unfolded domains, which would otherwise make them prone to inter-domain misfolding. Indeed, further studies with the first two domains of EF-G clearly demonstrate that both the ribosome and the molecular chaperone trigger factor suppress inter-domain misfolding (chapter V).

Chapter III. General materials and method

III-1: Buffers and Media

III-1.1: Buffers

5X SDS-PAGE sample buffer	0.25 M Tris-HCl pH 6.8
	0.5 M DTT
	10 % SDS
	50 % Glycerol
	0.5 % bromphenol blue
SDS-PAGE electrophoresis buffer	50 mM Tris-HCl pH 8.3
	380 mM glycine
	0.1% (w/v) SDS
5X running buffer for (Bis-Tris gel)	250 mM MES (26.65g)
	250 mM Tris-Base (15.14g)
	5 mM EDTA (0.931g)
	0.5% SDS (25mL of 10% SDS stock)
	pH 7.3
50X TAE-buffer	242 g/l Tris base
	57.1 ml/l acetic acid
	50 mM EDTA

10X TBE-buffer	108 g/l Tris Base
	55 g/l Boric Acid
	40 ml 0.5 M EDTA
	pH 8.0
TBST	25 mM Tris-HCl, pH 7.2
	150 mM NaCl
	0.1% Tween 20
TEV Storage buffer	20 mM Tris-HCl pH 8.0
	100 mM NaCl
	1 mM DTT
	0.5 mM EDTA
	10% glycerol
HN	25 mM HEPES pH 7.4
	100 mM NaCl
PBS	137 mM NaCl
	2.68 mM KCl
	10.1 mM Na_2HPO_4

	1.76 mM NaH_2PO_4
	pH adjusted to 7.4 with HCl
Biomix	13 mM Tris pH 7.0
	500 µM D-biotin
	100 mM ATP
	100 mM Mg-acetate
HKMß	20 mM HEPES pH 7.4
	100 mM KCl
	5 mM $MgCl_2$
	5 mM ß-ME
10X polymix-M buffer (no Mg^{2+})	5 mM $CaCl_2$
	50 mM NH_4Cl
	950 mM KCl
	10 mM spermindine-3HCl
	80 mM putrescin-2HCl

III-1.2: Media

LB medium	10 g/l bacto tryptone
	5 g/l bacto yeast extract

	5 g/l NaCl pH adjusted to 7.4 with NaOH
LB agar	16 g/l bacto agar dissolved in LB medium
SOC medium	20 g/l bacto tryptone 5 g/l bacto yeast extract 5 g/l NaCl 2.5 mM KCl pH adjusted to 7.4 with NaOH (after autoclaving, supplemented with) 10 mM MgCl2 20 mM glucose

III-2: DNA manipulations

III-2.1: General procedures

Gel extraction of DNA fragments, purification of PCR products and clean-up of digested DNA fragments were achieved by immobilization of DNA on a silica-based membrane, removal of contaminants and subsequent elution in 10 mM Tris-HCl, pH 8 (elution buffer, EB) or H_2O using commercially available kits (Wizard SV Gel and PCR Clean-Up System, Promega). Plasmid DNA was isolated from E. coli DH5α using the same principle (Wizard SV Miniprep

System, Promega). Double stranded-DNA was quantitated spectrophotometrically using NanoDrop (ThermoFisher Scientific) assuming that 1 A_{260} unit corresponds to 50 µg double stranded DNA. Endonucleases were purchased from New England Biolabs (NEB). Unless stated otherwise, Pfu turbo polymerase (NEB) was used to amplify DNA fragments with specific primers. PCR reactions were set up in the buffer supplied by the manufacturer without additional Mg^{2+} and contained 10 pmol of each primer, 10 pmol of each dNTP and 2.5 units of the polymerase in a volume of 50 µl. Denaturation during the PCR cycle was at 95 °C, annealing was at 50 °C and the elongation temperature was 72 °C unless otherwise indicated. After an initial denaturation of 2 min, 34 cycles of 30 s denaturation, 30 s annealing and elongation for a variable time were conducted. Plasmid was generally generated by Gibson Assembly from NEB. A colony of DH5α (NEB) cells transformed with the plasmid of interest was picked with a sterile pipet tip and first swirled in the PCR reaction mix before used to inoculate LB medium. Transformations were performed exactly following the corresponding competent cell protocols. Agarose gels were prepared in TAE buffer containing 1 – 2% agarose and 1:10000 dilution of SYBR Safe (Invitrogen) was added for visualization of double stranded DNA by blue LED light transilluminator (ThermoFisher Scientific).

The integrity of the open reading frame (ORF) of all plasmids was verified by sequencing (GENEWIZ). For bacterial growth, LB medium was used. When indicated, the medium was supplemented with 100 µg/ml ampicillin (LBAmp), 30 µg/ml kanamycin (LBKan), 170 µg/ml chloramphenicol (LBCam) or 50 µg/ml tetracycline (LBTet). Chemically competent E. coli DH5α and BL21(DE3) cells were purchased as indicated either from NEB or Agilent. For transformations of plasmid DNA, they were done exactly following the commercially available

protocols. Prior to plating on LB-agar, cells were allowed to recover at 37 °C for 1 h. For the preparation of electro-competent E. coli cells, a culture was grown in 5 ml LB-medium or SOC medium to an OD_{600} of 0.6 − 0.8. Cells were harvested and subjected to three successive washes with sterile water. Cells were resuspended in water, mixed with DNA and incubated on ice for 15 min.

Vector maps of all plasmids generated are given in the Appendices.

III-2.2: Cloning of EF-G constructs for purification and in-vitro translation

NEBbuilder (https://nebuilder.neb.com) was used to design the primers for backbone and inserts PCR. To generate an expression construct for full-length EF-G, I amplified the open reading frame from E. coli MC4100 genomic DNA using polymerase chain reaction and inserted it into pBAD His6 Sumo TEV LIC cloning vector (Addgene Plasmid #37507) that had been engineered to encode an N-terminal Avi tag (Beckett et al., 2008) and a C-terminal ybbR tag (Yin et al., 2006). After obtaining both the gel purified backbone and inserts PCR products, standard Gibson Assembly reactions were conducted following the commercial protocol with the commercial Gibson Assembly Master Mix kit (NEB #E2611s).

III-2.3: Immobilization of DNA to polystyrene beads

To functionalize polystyrene microspheres with DNA, I mixed either single-stranded or double-stranded oligonucleotides containing a 5’ amino-modifier with 2.1 μm carboxyl-polystyrene beads (Spherotech). Coupling of the oligonucleotide to the bead surface was achieved by chemical crosslinking with 1-ethyl-3- 3-dimethylaminopropyl] carbodiimide hydrochloride (EDC; ThermoFisher) in the presence of N-hydroxysulfosuccinimide (Sulfo-NHS) in 100 mM MES-Na (pH 6) buffer. The reaction was quenched by addition of TE8 buffer (20

mM Tris-HCl, 1 mM EDTA, pH 8). Unreacted material was removed by washing the beads three times with TE8. Beads modified with single-stranded DNA ("ssDNA-beads") were used to generate long DNA handles. Beads modified with double-stranded DNA ("dsDNA-beads") were used to immobilize proteins and RNCs. In this case, the double-stranded oligonucleotides contained a 4-nucleotide single-stranded overhang with a sequence reverse-complementary to the overhang of the CoA-modified double-stranded DNA (dsOligo-CoA DNA).

III-2.4: On bead PCR

To generate bead-immobilized DNA handles, the oligonucleotide on ssDNA-beads was extended by Taq polymerase. Beads were incubated with a linear DNA template that had the same sequence as the desired DNA handle, a 5'-biotinylated oligonucleotide that served as a reverse primer, and PCR mixture (containing Taq polymerase (New England Biolabs), manufacturer-provided buffer, and dNTPs). The reaction was subjected to thermal cycling, resulting in the synthesis of a 1789 bp long dsDNA handle on the bead-immobilized primer with a single biotin at the distal end. Beads were washed to remove unincorporated components and incubated with a large excess of streptavidin (ThermoFisher Scientific). Unbound streptavidin was removed by repeated washing of the beads. At the end of the procedure, "DNA handle-beads" contained covalently immobilized DNA handles (1789 bp long) which were stably bound at the distal end to a streptavidin molecule.

III-2.5: On bead ligation

To attach protein/RNCs of interest to polystyrene bead for optical tweezers measurement, I performed a on bead DNA ligation. The materials include pre-cleaned 2 µl of ~0.3% dsDNA-beads (~7 *10^5 beads/µl), 1.5 µl of 10 X ligation buffer (NEB B0202S), 1.5 µl ATP/Mg (10 mM

ATP, 10 mM Mg), 7 µl RNCs/protein sample and 3 µl of T4 ligase (NEB M0202S). The reaction was incubated at room temperature for 2 hours, and can be quenched by adding 2 µl of 50 mM EDTA for protein samples.

III-3: Protein preparative methods

III-3.1: Purification of EF-G constructs

Plasmids were transformed into BL21-Gold(DE3) (Agilent Technologies) host cells for recombinant gene expression. Expression and purification procedures were the same for all constructs including full-length EF-G, G_{alone}, *G-II*, *G-II-III*. Protein expression was induced with final concentration of 0.2% (w/v) L-Arabinose (AMRESCO) at 25°C when OD_{600} reached 0.4~0.6. Cells were harvested 5 hours after induction at 25°C (or 2 hours at 37 °C). Cells were lysed using an EmulsiFlex-C5 (Avestin) in buffer HN (25 mM HEPES-KOH, pH 7.4 100 mM NaCl). After cell lysis and centrifugation at 30,000 g, 4°C for half an hour to remove cell debris, the proteins were affinity purified from the supernatant using a 5 ml HisTrap column (GE Healthcare). The protein was dialyzed against buffer HN overnight, together with 1:1000 (w/w) Ulp1 to remove the His6-SUMO tag (Mossessova and Lima, 2000). The cleaved protein was applied to the HisTrap column again (after removal of imidazole by dialysis) to remove the His6-SUMO moiety and the His6-tagged Ulp1 enzyme. The flow through was saved for further modification. Purified protein was then incubated with BirA biotin ligase (1 µM) in 1X biotinylation buffer (500 µM D-biotin, 5 mM ATP, and 5 mM Mg-acetate) at 25°C for 1 hour to ensure complete biotinylation of the Avi-tagged protein. The product was concentrated and loaded onto a HiPrep 16/60 Sephacryl S-300 HR column (GE Healthcare) and eluted with

HKMD buffer (25 mM HEPES-KOH, 100 mM KCl, 5 mM Mg-acetate, 2.5 mM DTT). After concentration, protein aliquots were flash-frozen and stored at -80°C.

III-3.2: Purification of Trigger factor (TF)

TF was overexpressed with a TEV protease cleavable N-terminal His6-tag from a pPROEX-HTa based plasmid (Kaiser et al., 2006) in BL21-Gold(DE3) (Agilent Technologies) cells at 37°C for 3 hours. The cells were lysed at in buffer PBSI20 (137 mM NaCl, 2.68 mM KCl, 10.1 mM Na_2HPO_4, 1.76 mM Na_2HPO_4, 20 mM Imidazole, pH 7.4) by three passages through an EmulsiFlex-C5 (Avestin) homogenizer at 15,000 psi. The supernatant after centrifugation at 30,000 g, 4°C for 30 minutes was recovered and applied to a HisTrap column (GE Healthcare) charged with Ni^{2+}. The eluate was dialyzed against PBS buffer with 2 mM DTT together with His6-tagged TEV protease (2 mg/ml, purified in-house) at a ratio of 100:1 (v/v) overnight at 4°C. After digestion, the sample was passed over a HisTrap column again. His6-tagged TEV protease, un-cleaved His6-tagged TF and the cleaved His6-peptide were retained on the column. The flow through, containing TF without the His6-tag, was collected, flash-frozen and stored at -80°C.

III-4: Derivatization of protein with DNA handles

To introduce DNA handles for tethering, I incubated the purified ybbR-tagged proteins with CoA-modified double-stranded DNA (dsOligo-CoA) and in-house purified Sfp enzyme (Yin et al., 2006) in HM buffer (50 mM HEPES-KOH, 10 mM $MgCl_2$, pH 7.4) with final concentration of ~50 µM proteins, 50 µM dsOligo-CoA, and 5 µM Sfp. The enzyme couples the CoA-moiety to a specific residue in the ybbR-tag, resulting in position-specific, covalent coupling of the oligonucleotide to the protein. The dsOligo-CoA contained a 4-nucleotide single-

stranded overhang at the end distal to the CoA moiety. The reaction products were subjected to a Superdex 200 10/30 GL column (GE Healthcare) in HKMD buffer to separate the protein-DNA handle products from unmodified protein, unreacted oligonucleotide and enzyme. Modified proteins were concentrated, flash-frozen and stored at -80°C.

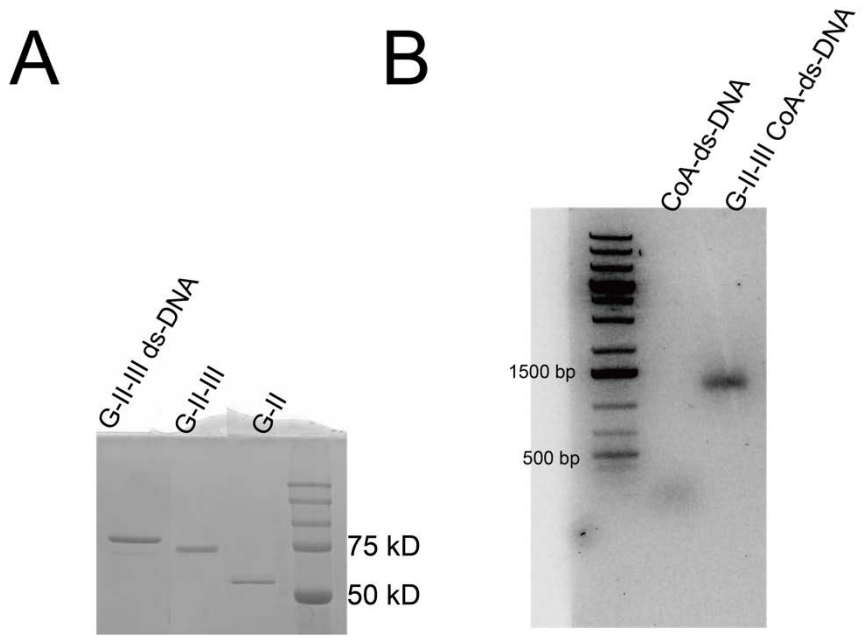


Figure III-1 Verification of derivatized *G-II-III* protein-DNA product. (A) SDS-PAGE of derivatized *G-II-III* protein-DNA product (lane 1), *G-II-III* without dsDNA modification (lane 2) and *G-II* (lane 3) to verify that the labeling of *G-II-III* with dsDNA was successful. (B) 1% TAE agrose gel to verify the *G-II-III* ds-DNA product. On the left is just CoA-ds-DNA, while on the right, the ds-DNA band is shifted up due to conjugation to the *G-II-III* protein.

III-5: Ribosome-nascent chain complexes (RNCs) preparation

RNCs were prepared as described before (Kaiser et al., 2011). Briefly, *in vitro* translation reactions were carried out using the PURExpress ΔRibosome Kit (New England Biolabs) that I supplemented with the purified ribosomes (~ 1 μM) that I had modified with dsOligo-CoA as described (Kaiser et al., 2011). Biotinylation of nascent polypeptides was conducted concomitantly with *in vitro* translation by the addition of biotin, ATP and BirA. Reactions were chilled on ice and briefly clarified by centrifugation (10 min, 16,000g, 4°C). RNCs were

sedimented through a 1 M sucrose cushion in buffer HKMß (25 mM HEPES-KOH, pH7.4, 150 mM KCl, 5 mM Mg-acetate, 5 mM β-mercaptoethanol) by ultracentrifugation for 40 min at 200,000 g, 4°C. The supernatant was discarded and the pelleted RNCs were resuspended in HKM buffer. Translation products were analyzed by SDS-PAGE followed by detection with Streptavidin-HRP (Southern Biotech) after electro-blotting onto a nitrocellulose membrane to confirm accumulation of the expected product.

RNCs were prepared as described previously (Kaiser et al., 2011). I generated PCR products that share the same 5' T7 promoter and ribosome binding site, but end at the indicated codon positions. The PCR products did not encode a stop codon. PCR products were i*n vitro* transcribed (T7 MegaScript Kit, Ambion), and the mRNAs were isolated (MegaClear Kit, Ambion) after digestion of template DNA. Stalled RNCs were generated using a reconstituted *in vitro* translation system (PURExpress ΔRibosome Kit, New England Biolabs) supplemented with ribosomes (Kaiser et al., 2011; Liu et al., 2017). Prior to the translation reaction, ribosomes were reacted with CoA-modified oligonucleotides and isolated by centrifugation. I utilized a >3-fold molar excess of mRNA over ribosomes to favor the formation of monosomes. The translation reaction also contained biotin and the BirA enzyme, resulting in biotinylation of the nascent polypeptide at an N-terminal Avi tag, which provide an attachment point in optical tweezers experiments. Because the mRNAs did not contain a stop codon, the translation product accumulated as a stably stalled peptidyl-tRNA that was sensitive to puromycin treatment, as assessed by Western blotting (Figure III-2). These RNCs were isolated by centrifugation through a sucrose cushion, dissolved in buffer HKMß, flash-frozen and stored at -80°C. Immediately

prior to optical tweezers experiments, RNCs were coupled to dsDNA-beads and tethered as described above.

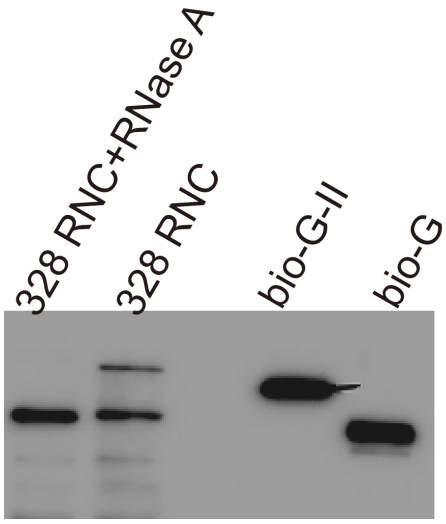


Figure III-2 Verification of 328_{RNC}. Proteins were electro-blotted onto a nitrocellulose membrane and detected with streptavidin-HRP. Lane 1: one band showing 328 amino acid protein size, the peptidyl-tRNA was digested by RNase A; lane 2: without RNase A digestion, the peptidyl-tRNA conjugated 328_{RNC} (G-RNC, top band) appeared at a higher molecular weight; lane3: empty; lane 4: isolated biotinylated *G-II* protein, which should have higher molecular weight than 328-amino acid protein size; lane 5: isolated biotinylated G-domain, which should have lower molecular weight than 328-amino acid protein size.

III-6: Single-molecule optical tweezers experiments

III-6.1: Optical tweezers force ramp experiments

I carried out optical tweezers experiments using a home-built instrument with a single trap formed by counter-propagating beams of two 845 nm diode lasers (Smith et al., 2003). All optical tweezers experiments were performed in buffer HKMß. One 2.1 μm polystyrene bead with protein of interest was immobilized by the micropipette through suction, and the other bead with only DNA handle was held by the optical trap. Single tethers were confirmed either through DNA overstretching (Smith et al., 1996) or the signature unfolding transition from G-domain. All force ramp data was collected with pulling velocity of 100 nm/s, a trap stiffness of around 0.1 pN/nm, and a sampling rate of 1000 Hz. The minimum force was set to 2 pN if not otherwise

specified, the maximum force as well as folding time at minimum force vary among experiments. A full cycle of force ramp experiment is defined as pulling from the minimum to the maximum force, relaxing back to the minimum force and waiting time at the minimum force for the indicated time (Figure III-3A). The cycles were repeated until the tether broke.

III-6.2: Optical tweezers force clamp experiments

Prior to force-clamp measurements, I subjected each tethered molecule to several force ramp cycles to verify that it exhibited the expected characteristic unfolding signature. In force clamp measurements, tethered proteins or RNCs were first stretched to high force until unfolding was observed. Subsequently, the force was set to low force in one step. While holding the molecule at this force, changes in molecular extension over time were followed to monitor the folding process of the molecules at a sampling frequency of 1000 Hz. After refolding of the protein, the process of unfolding at high force and refolding at low force was repeated.

III-7: Single-molecule data analysis

III-7.1: *Constant velocity data analysis*

During constant velocity (force ramp) experiments, the optical trap was continuously moved, resulting in a continuously increasing force on the tethered molecule (Figure III-3A). The unfolding rate increases exponentially with the applied force and therefore continuously increases in this type of experiment until the molecule ultimately unfolds. Unfolding transitions are apparent as “rips” in the force extension curves (FECs). Each rip is characterized by an unfolding force (F_{unf}) and a change in molecular extension (Δx_{unf}). As the protein unfolds, it transits from a stiff, compactly folded structure to a soft, extended polypeptide. As a

consequence, the stiffness of the tether decreases, while the contour length increases simultaneously, causing a drop in force. Subsequently, the force increases again because the trap continues to move.

The F_{unf} was determined by locating discontinuities in the FECs using a custom Matlab script. To determine Δx_{unf}, the difference in extension before and after the rip at F_{unf} was measured (Figure III-3B). Because the extensions are measured at the same force, parts of the tether that do not change (e.g. the ribosome, domains that remain folded, DNA handles) have no effect on the measured extension change. The measured change therefore reflects only the contribution from the part of the polypeptide that has unfolded in a particular transition. The measured extension change is a relative change, i.e. it is independent of the absolute length of the tether. To better characterize the unfolded protein, the interpolation formula of worm like chain model was used:

$$F(x)_{wlc} = \frac{k_b T}{l_p}\left(\frac{1}{4}\left(1 - \frac{x}{L_c}\right)^{-2} - \frac{1}{4} + \frac{x}{L_c}\right),$$

where F is the force, k_b is the Boltzmann constant, T is the absolute temperature, l_p is the persistence length of protein, which is 0.65 nm if not otherwise specified, x is the extension that was observed in force ramp experiment, L_c is the contour length, which is 0.36 nm * N_{aa}, where N_{aa} is the number of amino acids of the protein.

Because the lifetimes of a given folded structure at any given force are stochastic, unfolding does not occur at a single discrete force in single-molecule force ramp experiments. Instead, the unfolding forces for a particular structure follow a characteristic distribution (Dudko et al., 2008) (Figure III-3C). The unfolding force distribution can be expressed as:

$$P(F) = \frac{\exp\left(-\int_0^F [\dot{F}(f)\tau(f)]^{-1} df\right)}{\dot{F}(F)\tau(F)},$$

Where P(F) is the probability of unfolding force at F, $\dot{F}(f)$ is the force loading rate at given force, $\tau(F)$ is the life time at given force. The procedures of numerically transforming the unfolding force distribution to force dependent rates were as following:

- Bin the unfolding force into histograms, which contains N bins of width ΔF that starts at F_0 and ends at $F_N = F_0 + N\Delta F$. If the counts for each bin is C_i at the *i*th bin, then the probability of F is $P(F) = C_i/(N_{tot}\Delta F)$, where N_{tot} is the total number of unfolding forces.
- Then the life time of the *k*th bin of force can be written as:

$$\tau\left(F_0 + \left(k - \frac{1}{2}\right)\Delta F\right) = \frac{\left(\frac{h_k}{2} + \sum_{i=k+1} h_i\right)\Delta F}{h_k \dot{F}\left(F_0 + \left(k - \frac{1}{2}\right)\Delta F\right)}.$$

To extrapolate the zero-force life time, the simplest Bell-Evans model was adopted (Bell, 1978) (Figure III-3D), which can be written as:

$$\tau(F) = \tau_0 \exp\left(\frac{-Fx^{\ddagger}}{k_b T}\right),$$

where $x^{\ddagger}$ is the distance to transition state. Of course, it is also well known that the distance to the transition state could be force dependent, so a more sophisticated model that takes into account of both the barrier height shape of energy landscape can be used (Dudko et al., 2006). A potential problem with the latter model could be overfitting. Therefore, experiments with more than one loading rates are normally performed and multiple shapes of the energy landscapes should be tested.

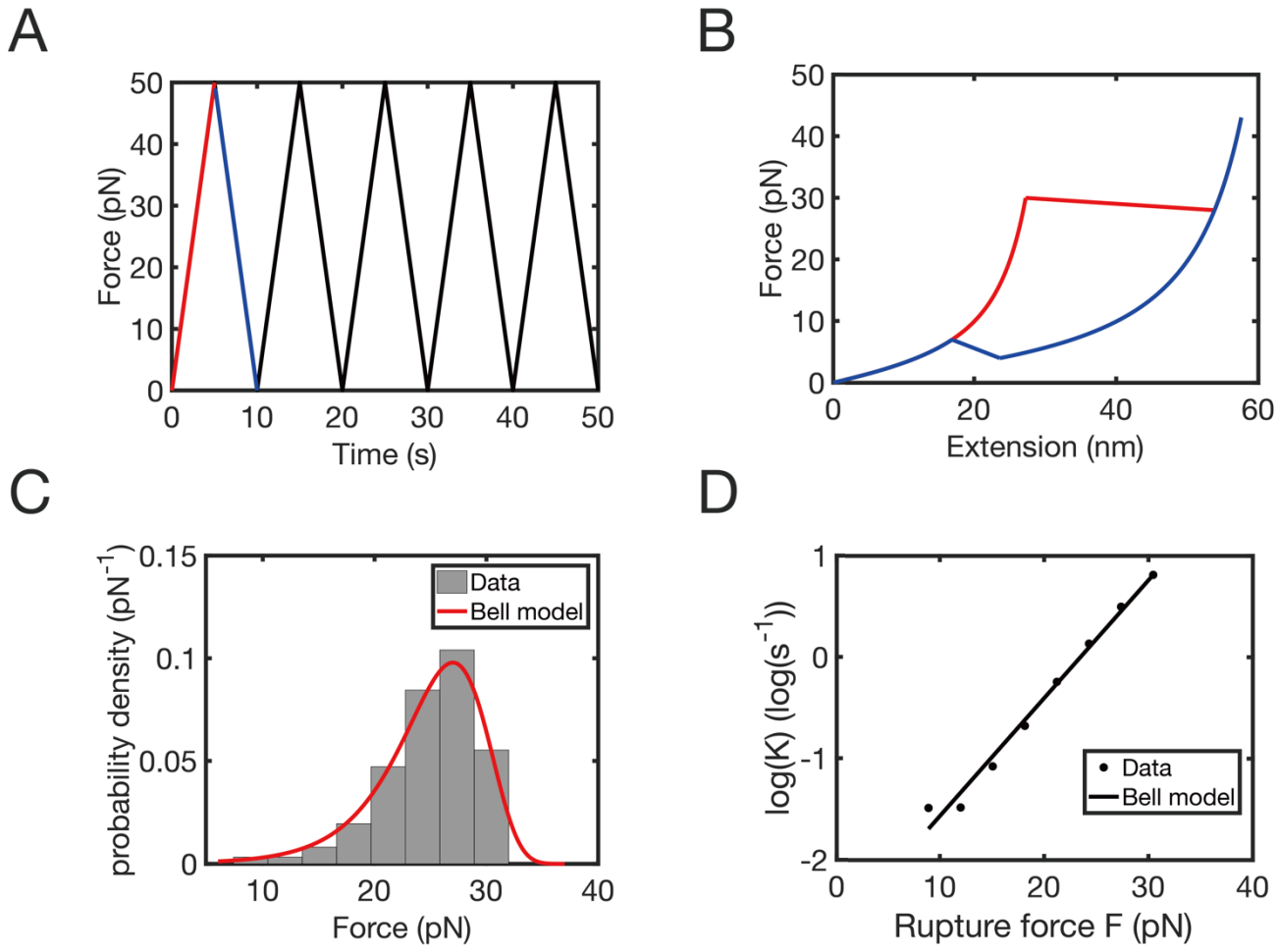


Figure III-3 Constant velocity experiment and data analysis. (A) Example of force vs time plot from a hypothetical constant velocity experiment. The initial stretching is plotted in red line, and relaxation is plotted in blue line. The rest ones are plotted in black. (B) Example of force vs extension plot from the same hypothetical experiment in A, with only the initial stretching (red) and relaxation (blue) curves. The rips in red and blue curves indicate unfolding and refolding of the same molecule. (C) Unfolding force distribution from a hypothetical one step unfolding protein. The red curve is a fit of the force distribution using Dudko's method combined with Bell's model. (D) Transformation of unfolding force distribution to force dependent rates using Dudko's method. The rates are plotted in log scale, and the black line is a fit using Bell's model.

III-7.2: *Constant force data analysis*

Force ramp experiments are usually non-equilibrium experiments, meaning that hysteresis is commonly observed in force ramp experiments. However, these experiments allow obtaining transitions easier without too much prior knowledge about the system. After acquiring enough information about the system, for example, knowing what the average unfolding and

refolding forces in force ramp experiment are, near equilibrium measurement at certain constant force can be performed. A typical trajectory of a single protein folding and unfolding in equilibrium is shown as Figure III-4A. The unfolded and folded states are usually distinguished by differences in either extensions or kinetics. In order to assign the observed states, Hidden Markov Model (HMM) is usually used. It uses probabilistic arguments to associate the individual data points of trajectory (o_1, o_2, ……, o_T), to hidden states x_i (Figure III-4B). The hidden states and their transition probabilities are defined by a Markov process and each hidden state with a certain emission probability to create a certain observable. The forward-backward algorithm allows the determination of the most probable state of the hidden network for each point of the trajectory. Using them together with Bayes' rule (Chodera et al., 2011) allows the determination of the probabilities P to be in the hidden states x_i at a certain time point t for all measured observables:

$$P\left(x_i \middle| o_{1,} o_{2,} \dots, o_T\right) \propto \alpha_t\left(i, o_{1,} o_{2,} \dots, o_T\right)\beta_t\left(i, o_{1,} o_{2,} \dots, o_T\right),$$

where α_t is the forwards variable, while β_t is the backwards variable. After assigning the states, as in the case of Figure III-4A, the life time distribution can then be fit by single exponential distribution to extract the mean life time $\langle\tau\rangle$ (Figure III-4C), which is also 1/k, where k is the unfolding/folding rate. In more complex cases, where there are multiple states, the rate can then be written as:

$$k_{ij} = -\ln\left(1 - T_{ij}\right) \cdot \delta,$$

where k_{ij} is the transition rate from state i to j, T_{ij} is the transition probability from state i to j, and δ is the sampling rate.

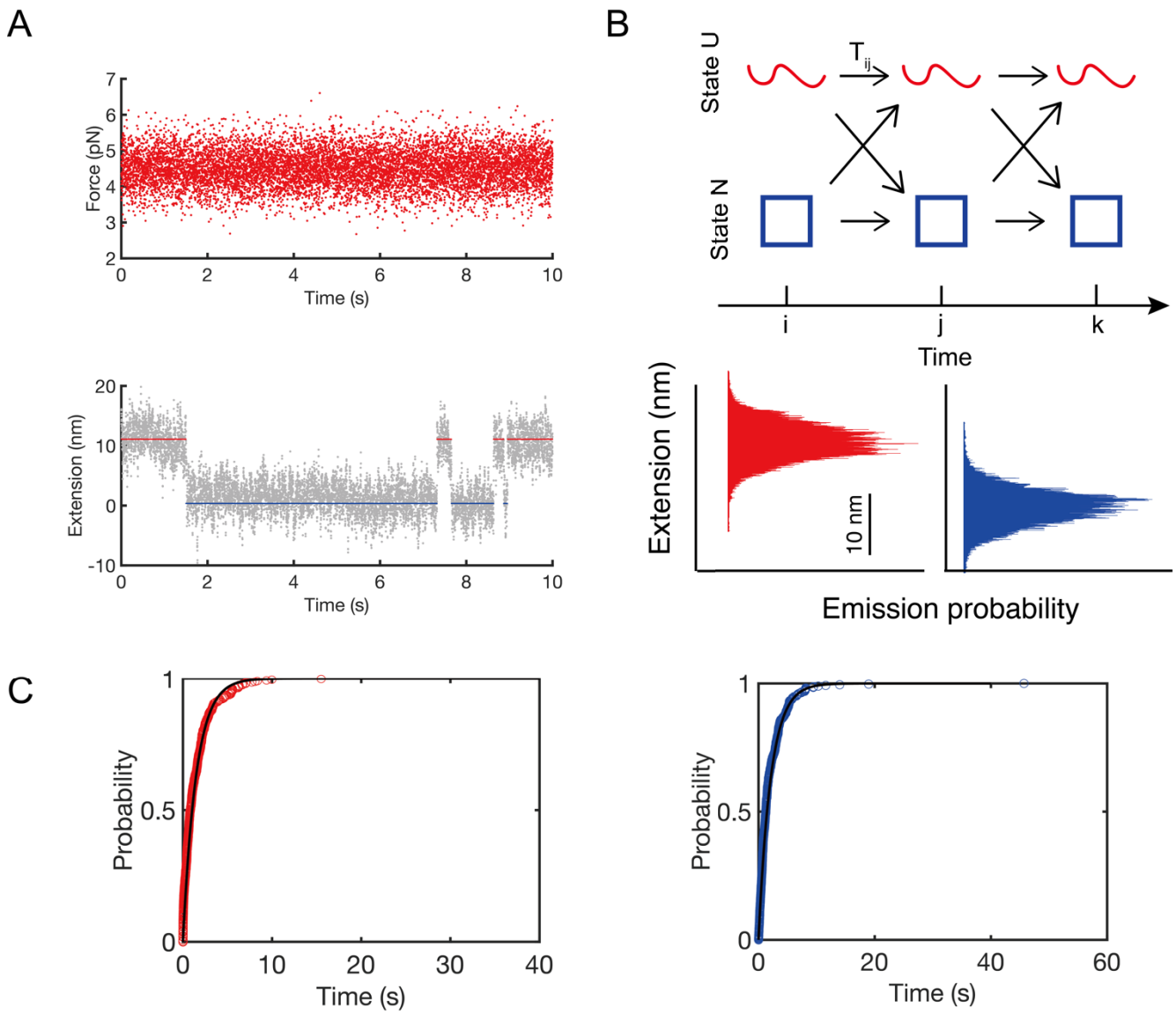


Figure III-4 Constant force experiment and data analysis. (A) A hypothetical constant force experiment example. Top: constant force vs time plot, the force fluctuates at around 4.5 pN. Bottom: Extension trajectory, at the applied force (top). The extension changes over time, hopping back and forth between two states. The higher extensions are the unfolded state (marked by red lines), and the lower extensions are folded state (marked by blue lines). (B) Hidden Markov model (HMM) analysis for the extension trajectory. Using HMM allows determining the states of the trajectory as well as the transition probabilities between states. The emission probabilities can be estimated by analyzing the extension distributions of the states. In order to obtain the optimal model and accurate parameters estimation, maximum likelihood-based methods or Bayesian statistics combined with HMM are most commonly used. (C) Dwell time distribution of unfolded state, which can be used to estimate the folding rate. (D) Dwell time distribution of folded state, which can be used to estimate the unfolding rate.

.

Chapter IV: The ribosome destabilizes a nascent multidomain protein

By

Christian Kaiser and Kaixian Liu

This work has been published as:

Liu, K., Rehfus, J.E., Mattson, E., and Kaiser, C.M.. (2017). The ribosome destabilizes native and non-native structures in a nascent multi-domain protein. *Protein Sci.* 26, 1439-1451.

Contributions: Kaixian Liu conducted most of the experiments. Joseph Rehfus and Elliote Mattson performed some of the G-RNC force ramp experiments. Kaixian Liu and Christian Kaiser analyzed the data, created figures, and wrote the manuscript.

Chapter IV. The ribosome destabilizes a nascent multidomain protein

IV-1: Introduction

Combining multiple domains into a single polypeptide chain is a widely used evolutionary strategy for generating proteins with novel functions (Han et al., 2007). Consequently, all proteomes contain a substantial fraction of multidomain proteins. Like all cellular proteins, these large proteins are synthesized as linear polypeptides by the ribosome and must fold into precisely ordered structures to become biologically active. While small proteins, which often fold rapidly into their native structures, have been extensively studied (Dill and MacCallum, 2012; Sosnick and Barrick, 2011), we know much less about the principles that govern the folding of large, multidomain proteins (Braselmann et al., 2013).

Large proteins often fail to refold efficiently upon dilution from denaturant and instead form insoluble aggregates of misfolded species, making it challenging to study their folding experimentally (Braselmann et al., 2013). In vivo, however, these proteins fold faithfully to avoid misfolding and potentially toxic aggregation (Dobson, 2003), even in the crowded cellular environment. Thus, the traditional approaches that have provided invaluable insight into the biophysical principles of folding are not capturing features that are crucial for complex, multidomain proteins.

In the cell, proteins begin to fold while they are still being synthesized by the ribosome. Co-translational folding (Rodnina and Wintermeyer, 2016) is a likely mechanism to reduce the complexity of folding by allowing sequential, domain-wise folding. Nascent proteins remain in close proximity to the ribosomal surface and are not released until their synthesis is complete.

The ribosomal environment and the process of synthesis itself can have profound consequences on the folding efficiency and, in some cases, the folding outcome (Cabrita et al., 2016; Evans et al., 2008; Hoffmann et al., 2012; Kaiser et al., 2011; Khushoo et al., 2011; Kimchi-Sarfaty et al., 2007; Nicola et al., 1999; Siller et al., 2010; Spencer et al., 2012; Ugrinov and Clark, 2010; Zhang et al., 2009). Several studies have found that interactions with the ribosome destabilize nascent proteins (Cabrita et al., 2016; Kaiser et al., 2011; Knight et al., 2013; Samelson et al., 2016) and reduce their folding rates (Kaiser et al., 2011), but the underlying mechanisms are not well understood. These studies were largely carried out using robustly folding small proteins or domains that were tethered to the ribosome through an artificial C-terminal extension. How the ribosome affects multidomain protein folding has remained largely unexplored.

Here, I have begun to develop elongation factor G (EF-G) from *Escherichia coli*, a GTPase with an essential function in translation elongation, as a model to study how the ribosome affects the folding of a cytosolic multidomain protein. EF-G has five domains (termed G-domain, II, III, IV, and V, Figure IV-1) that comprise a total of 704 amino acids (Pulk and Cate, 2013). All organisms contain EF-G homologs. The combination of an N-terminal GTP-binding domain (G-domain) with additional domains, as in EF-G, represents an even more widespread motif (Yamamoto et al., 2014). Thus, the EF-G structure represents a common scaffold, variations of which give rise to a number of distinct proteins (Yamamoto et al., 2014). Understanding the folding of the complex multidomain protein EF-G may thus shed light onto the properties that guide the folding of a large group of proteins that are built on similar structural scaffolds, and enable comparisons among these proteins to determine to what degree folding pathways are conserved among homologs (see appendices).

I carried out single-molecule experiments with optical tweezers to characterize the folding of full-length EF-G and of the N-terminal G-domain, both in isolation and as a ribosome-bound nascent polypeptide. Full-length EF-G refolds very inefficiently due to interactions among unfolded domains that are detrimental to productive folding. The G-domain alone can fold cotranslationally, indicated by its ability to adopt stable structures in isolation and on the ribosome. Interestingly, the domain adopts both native and non-native structures, both of which are destabilized by the ribosome. This destabilization may help the G-domain to avoid permanently misfolded conformations and reach its native structure, completing the first step of EF-G folding.

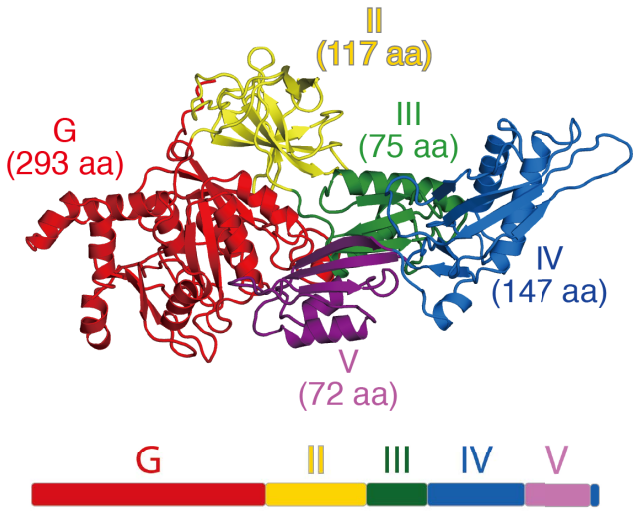


Figure IV-1 Structure diagram of EF-G. Ribbon diagram (top) of the *E. coli* EF-G crystal structure (pdb: 4V9P) and bar diagram (bottom) of the primary structure.

IV-2: Results

IV-2.1: Full-length EF-G unfolds domain-wise

To obtain insight into the stability of EF-G, I mechanically unfolded the full-length protein (Figure IV-1) using optical tweezers. I engineered a construct containing Avi (Beckett et

al., 2008) and ybbR (Yin et al., 2006) tags at both termini that enable the attachment to polystyrene beads for mechanical manipulation through DNA handles (see Figure IV-2). Applying mechanical force destabilizes folded structures, resulting in unfolding. Unfolding events are apparent as "rips" - sudden increases in the molecular extension that accompany the transition from a compactly folded to an unfolded structure that is extended under force.

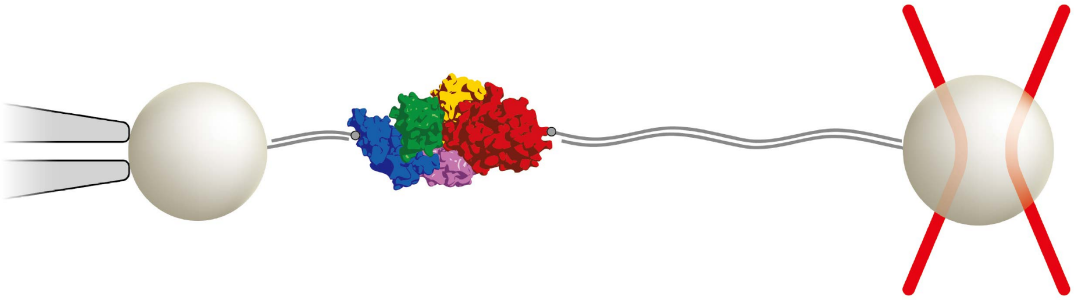


Figure IV-2 Optical tweezers experiment setup (not to scale). Schematic of the experiment setup for unfolding experiments using optical tweezers.

I tethered the purified protein between two polystyrene microspheres and subjected it to "force ramp" experiments. In these measurements, I continuously increased the applied force by moving the optical trap at a velocity of 100 nm/s. Initial stretching of the molecule resulted in a sequence of distinct unfolding transitions over a force range of 2 to ~45 pN, as illustrated by the representative force-extension curve shown in Figure IV-3A (red trace). I obtained similar results for 10 molecules. The extension changes and unfolding forces in the initial unfolding trace ("first pull") of each molecule are shown in Figure IV-3B. Combining the contour length changes of the transitions in each trace, calculated from the extension change at the unfolding force using the worm-like chain (WLC) model (Bustamante et al., 1994), yielded 246.0 nm ± 2.4 nm (standard deviation, std), close to the expected value of 244.5 nm for complete unfolding of natively folded EF-G based on the crystal structure (PDB code: 4v9p (Pulk and Cate, 2013)). This observation is

consistent with the notion that EF-G in our construct is natively folded and completely unfolds in the force range probed in our experiments (2 to 50 pN).

The step-wise unfolding over a large range of forces suggests that at least some of the EF-G domains unfold independently. I calculated the expected length changes for each domain based on the EF-G crystal structure (Pulk and Cate, 2013) (see Table IV-1).

domain	n_1	n_2	Δx_{native} (nm)	$\Delta L_{C,calc}$ (nm)
G	1	293	2.6	102.9
II	294	410	1.9	40.2
III	411	485	2.5	24.5
IV + V	486	699	1.5	75.5
V	615	686	0.6	25.3

Table IV-1 EF-G domain dimension calculations. N_1 is the starting amino acid residue position and n_2 is the ending amino acid position. Δx_{native} is the end to end distance of native state based on the crystal structure. ΔL_{Ccalc} is the expected contour length change after unfolding for each domain.

The transitions at the highest force, occurring between 30 and 45 pN, show a mean contour length change of 102.3 nm (± 2.2 nm, std), in good agreement with the 102.9 nm expected for unfolding of the N-terminal G-domain (Figure IV-3B, red line). Experiments with the isolated G-domain (see below) confirm the assignment to G-domain unfolding.

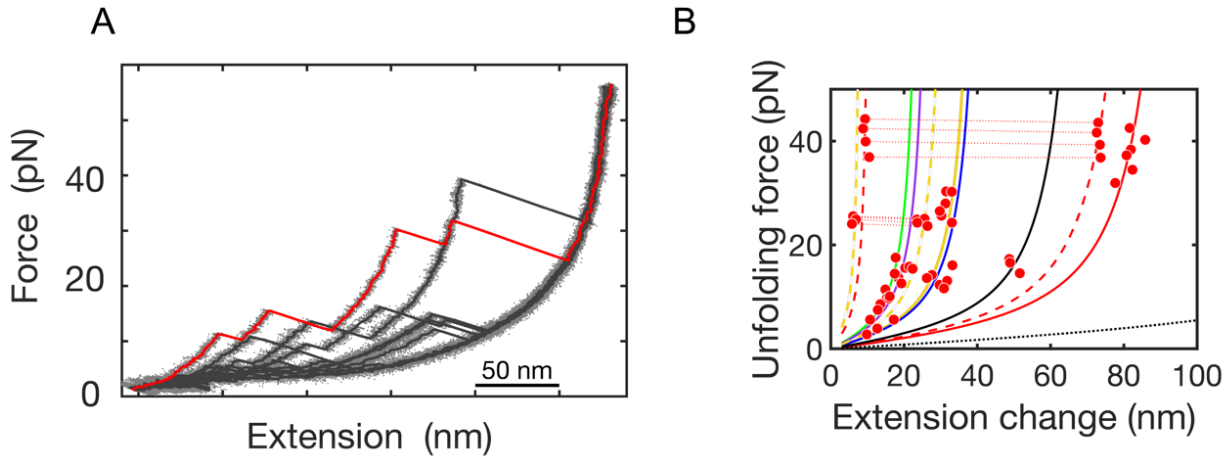


Figure IV-3 Full-length EF-G unfolds sequentially. (A) Representative set of FECs recorded with a single EF-G molecule (grey dots: raw data, 1000 Hz, lines: filtered to 33 Hz). The initial extension of the molecule is drawn in red and shows distinct transitions for the five domains. Traces recorded after relaxing the molecule and allowing refolding to proceed for 10 seconds are shown in black (relaxation is not plotted). None of the traces after refolding shows the complete set of transitions apparent in the first curve, but individual domains occasionally refold. (B) Scatter plot showing unfolding transitions during the initial extension. The events match the extension changes expected based on the EF-G structure (indicated by WLC curves). Solid lines indicate WLC curves that match full domains, while dashed lines indicate individual steps in two-step transitions. The line colors are the same as the corresponding domain colors in Figure IV-1. The G-domain and domain II populate unfolding intermediates. Dotted red lines connect events that occur in quick succession in a single trace (with corresponding WLC curves as dashed lines). Domain IV and V sometimes unfold in one step (black line).

Transitions in the range of 24 to 30 pN show extension changes of 41.7 nm (± 1.4 nm, std), similar to what is expected for unfolding of domain II (40.2 nm; yellow line in Figure IV-3B). Mechanical unfolding experiments with a construct encompassing the G-domain and domain II confirm that these transitions represent unfolding of domain II (Figure IV-4). The G-domain and domain II both unfold in either a single transition or two successive transitions that add up to the expected length change (dashed red and yellow lines in Figure IV-3B).

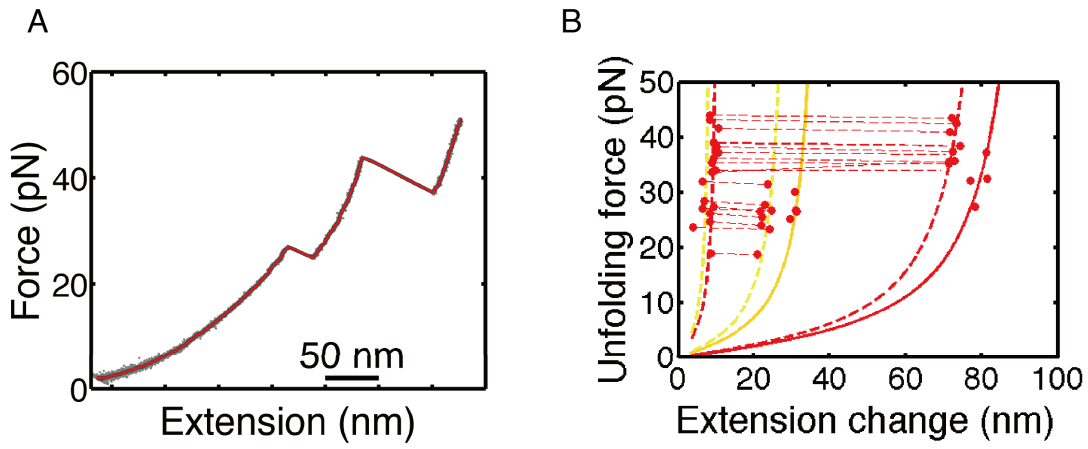


Figure IV-4 *G-II* unfolds sequentially. (A) Example FEC of initial unfolding of *G-II*, showing two transitions at different forces, one corresponding to domain II unfolding at lower force. (B) Scatter plot of initial unfolding transitions from *G-II*. The events match the transitions from full-length EF-G that were defined as domain II and G-domain unfolding. The WLC curves are plotted the same as Figure IV-3.

While the structure of the G-domain unfolding intermediate is unclear, the dimensions of the domain II unfolding intermediate are consistent with rupture of the G-domain/domain II interface, suggesting coupling between these two domains (chapter V). Domains III, IV and V unfold at forces below 20 pN. Domains III and V have very similar calculated extension changes associated with unfolding (24.5 nm for domain III, 25.3 nm for domain V), making it difficult to distinguish them based on length change alone. Unfolding events with similar length changes are observed in the force range of ~5 to 20 pN. A cluster of unfolding events in a single step below ~15 pN is observed in all traces. The ~25 nm transitions above that force are observed in only some traces. The remaining traces show a much longer transition of ~75 nm that agrees well with simultaneous unfolding of domains IV and V. Simultaneous unfolding of these two domains is readily rationalized by their organization in the primary structure (see Figure IV-1), which suggests coupling between these domains: Domain IV is discontinuous and straddles domain V. Further experiments with constructs *G-II-III* and *III-IV-V* confirmed our assignments of unfolding transitions in full-length EF-G to individual domains (Chapter VI). However, based on

the reasoning above, it is very likely that domain III unfolds first (green line in Figure IV-3B), followed by domains IV and V either individually (blue and purple lines) or in combination (dashed blue/purple line in Figure IV-3B). In summary, the observed extension changes largely match the expectation based on the EF-G crystal structure. Overall, our findings indicate that – with the exceptions pointed out above – the EF-G domains mostly unfold individually.

IV-2.2: EF-G refolds inefficiently

To probe the folding of full-length EF-G after complete unfolding, I relaxed the force on the polypeptide to initiate refolding and held the molecule at 2 pN for 10 seconds. Subsequent stretching typically yielded different sets of transitions (Figure IV-3A). After refolding, I never observed a subsequent force-extension curve that exhibited a complete sequence of unfolding transitions resembling the initial unfolding trace, indicating that the protein did not fold back completely into its initial structure within the given refolding time interval. Many of the force-extension curves recorded after the refolding pause exhibited poorly defined unfolding transitions at low forces ($F < 5$ pN) with length changes that could not be reliably measured and are thus not represented in Figure 1D. The transitions observed at low forces (between 5 and 10 pN) with variable extension changes perhaps represent collapsed, molten globule-like states without a high degree of tertiary structure. Some of these structures unfold with extension changes substantially longer than expected for the largest domain (i.e., to the right of the solid red line in Figure IV-5), indicating that sequences from several domains participate in their formation. Overall, the protein mostly adopts mechanically labile and heterogeneous structures, suggesting that interactions among unfolded domains interfere with productive folding.

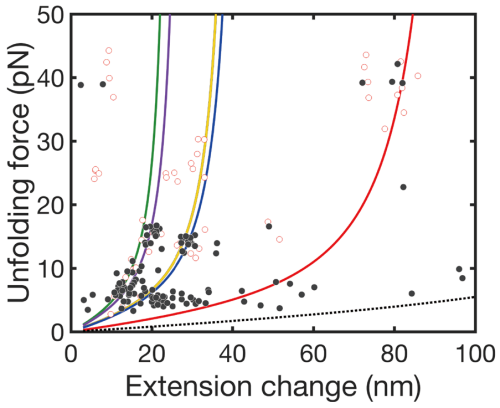


Figure IV-5 EF-G refolding is inefficient. Scatter plot of unfolding events observed after refolding (black dots). The transitions from Figure IV-3B are re-plotted as open circles for references. All domains except domain II show occasional refolding. The line colors are coded the same as in Figure IV-3.

While EF-G failed to completely fold back into its native structure within the 10 seconds refolding time, some of the transitions observed during initial unfolding are also apparent after the refolding pause (Figure IV-5), suggesting that some of the domains regain their native structures, although infrequently. Unfolding rips similar to those observed for domains III, IV and V are apparent in some of the force-extension curves, although they are partially obscured by the heterogeneous ensemble of transitions in the lower force range. The transition corresponding to G-domain unfolding is clearly observed with a probability of 0.15 (Figure IV-5). In some cases, the transition occurs in the same two-step manner as during initial unfolding, indicating that the G-domain refolds to its native structure in our experiments, although rarely. This observation indicates that the G-domain represents an independently folding unit that could acquire its native structure cotranslationally.

IV-2.3: The G-domain folds autonomously

Since the G-domain is produced first during EF-G synthesis and can fold into its native structure independently, its cotranslational folding could be particularly important for proper

structuring of more C-terminal domains. To obtain insight into the folding of the N-terminal G-domain, I generated a construct that comprises the N-terminal 294 amino acid residues of EF-G with tag sequences on either terminus that allow tethering for optical tweezers manipulations. I subjected the protein to force ramp measurements under the conditions described above and observed unfolding transitions (Figure IV-6A) that resembled those assigned to the G-domain in unfolding experiments with the full-length protein (Figure IV-3), indicating that the isolated G-domain represents an independently folding unit.

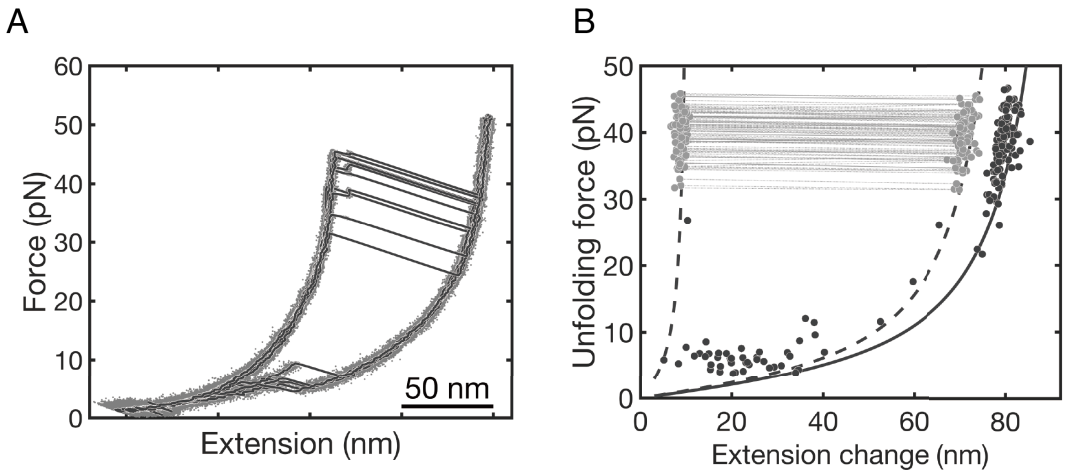


Figure IV-6 G-domain folds autonomously. (A) Representative set of successive FECs collected with a single G-domain molecule. The traces represent repeated extensions separated by a 10 s refolding pause. (B) Scatter plot representing the unfolding events seen in panel A. Events in which the unfolding intermediate could be resolved (two-step events) are plotted in grey with connecting lines; one-step events are plotted in black. The solid black line represents a WLC curve expected for one-step G-domain unfolding, dashed lines indicate WLC curves for two-step unfolding.

When the denatured protein was allowed to refold for 10 seconds at 2 pN, the G-domain refolded with a high probability (p_{refold}(10 s) = 0.71, 202 refolding events in 284 attempts), indicated by an extension change consistent with the native structure at high force (30 to 50 pN) (Figure IV-6A). When the molecule fails to refold, I either do not observe a measurable transition, or unfolding at forces below 10 pN, probably reflecting the unraveling of a loosely structured state that might correspond to an early molten globule-like folding intermediate.

Shortening the refolding pause to 1 second resulted in a decreased folding probability of p_{refold}(1 s) = 0.32, while increasing the waiting time at 2 pN to 15 seconds increased the folding probability to p_{refold}(15 s) = 0.79. A high likelihood of forming additional tethers between the beads during longer waiting times prevented us from extending the refolding time beyond 15 s. This increase in refolding probability over time indicates that the lower force unfolding events in our G-domain force ramp measurements represent partially folded states that will ultimately convert to the more stable structure, given enough time. Overall, the isolated G-domain robustly folds into mechanically stable structures that unfold in a force range of 30 pN to 50 pN.

IV-2.4: The G-domain transiently populates an unfolding intermediate

A fraction of the high-force unfolding transitions of the G-domain occurs in two distinct steps, whereas only a single step is resolved in the remaining unfolding transitions (Figure IV-6A and B). The single transitions show a contour length change of 101.4 nm (± 1.82 nm, std). For 2-step transitions, a small rip with a contour length change of 11.1 nm (± 1.03 nm, std) is followed by a larger transition of 89.9 nm (± 1.55 nm, std) within a time interval of 200 milliseconds, indicating that the protein populates an unfolding intermediate that is highly unstable in the unfolding force range. Combining the transitions in each trace yields a total change of 101.0 nm (± 1.52 nm, std), very close to the value obtained for 1-step transitions. The fraction of 2-step unfolding events among all unfolding events is $f_{2\text{-step}} = 0.39$. It thus is possible that either the molecule populates different states that yield similar total unfolding extension changes, or that the molecule unfolds through two different pathways. Alternatively, the short lifetime of the intermediate could result in failure to detect the intermediate at our sampling rate (1 kHz).

I first estimated the probability of failing to detect the unfolding intermediate due to the time resolution of our measurement. I only counted the intermediate if it was populated for at least 5 samples during the measurement, i.e. if it persisted for at least 5 milliseconds, because shorter dwell times do not permit reliable measurements of the change in molecular extension. The unfolding intermediate lifetime appeared to be constant over the unfolding force range (Figure IV-7B, inset). Determining the lifetime of the unfolding intermediate yielded a distribution that we fit to a single-exponential probability density function, yielding a mean lifetime of τ_{Iu} = 34.1 ms (with 95% confidence intervals, C.I., of 31.6 ms and 37.0 ms) (Figure IV-7B). By integrating the probability density function from 0 to 5 ms, we calculate the probability of missed detection of the intermediate to be p_{miss} = 0.14.

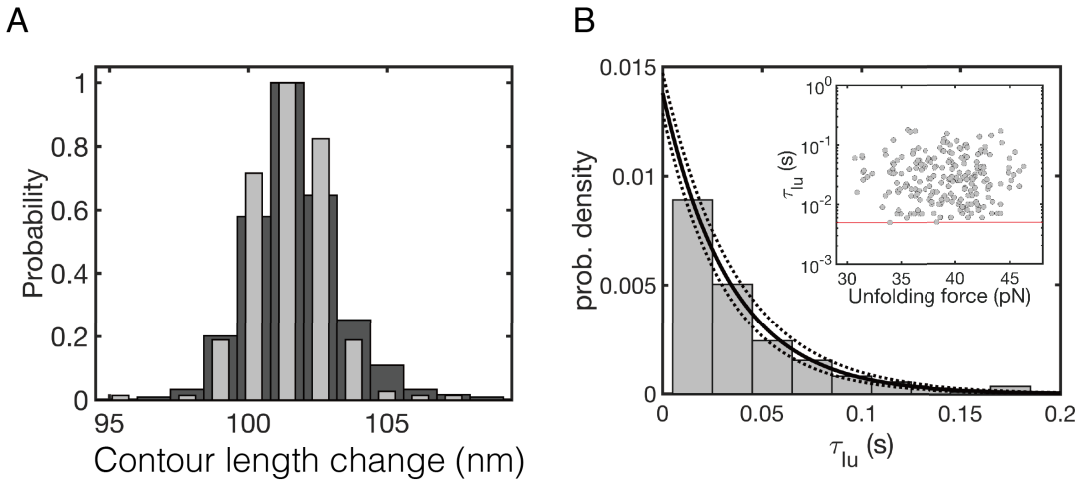


Figure IV-7 G-domain unfolds via an intermediate state. (A) Histogram of contour length changes associated with unfolding. Combined changes from two-step transitions are shown as light grey bars, one-step events are represented in black, showing similar distributions. (B) Histogram of the unfolding intermediate lifetime (τ_{Iu}) probability density (grey bars) with a single exponential fit (black lines; dotted lines represent 95% confidence intervals of the fit). The inset shows lifetimes of individual events at their respective forces. The red line represents the detection cutoff (5 ms) imposed by our sampling rate.

IV-2.5: The G-domain adopts distinct structured states

The above result makes it very unlikely that our finite sampling rate alone can account for the fraction of traces with only one transition ($f_{1\text{-step}} = 1 - f_{2\text{-step}} = 0.61$). Instead, at least two unfolding processes must underlie the observed behavior, one that populates the intermediate, and one that occurs in a single step. This notion is further supported by the observation that the 1-step and 2-step events occur with distinct unfolding force distributions (Figure IV-8A): the distribution of 1-step transitions is skewed toward lower values compared to that of 2-step transition. Thus, while the 2-step events represent a homogeneous population of unfolding events, the one-step events are likely a mixture of two components: one component represents events that are actually two-step transitions in which the intermediate has not been resolved in our measurements (see above), the other component represents a process with different unfolding parameters that, on average, occurs at lower force. Since we observe the unfolding intermediate in many of the initial unfolding curves as well as in the full-length protein (Figure IV-3), its population is likely a hallmark of the natively folded G-domain.

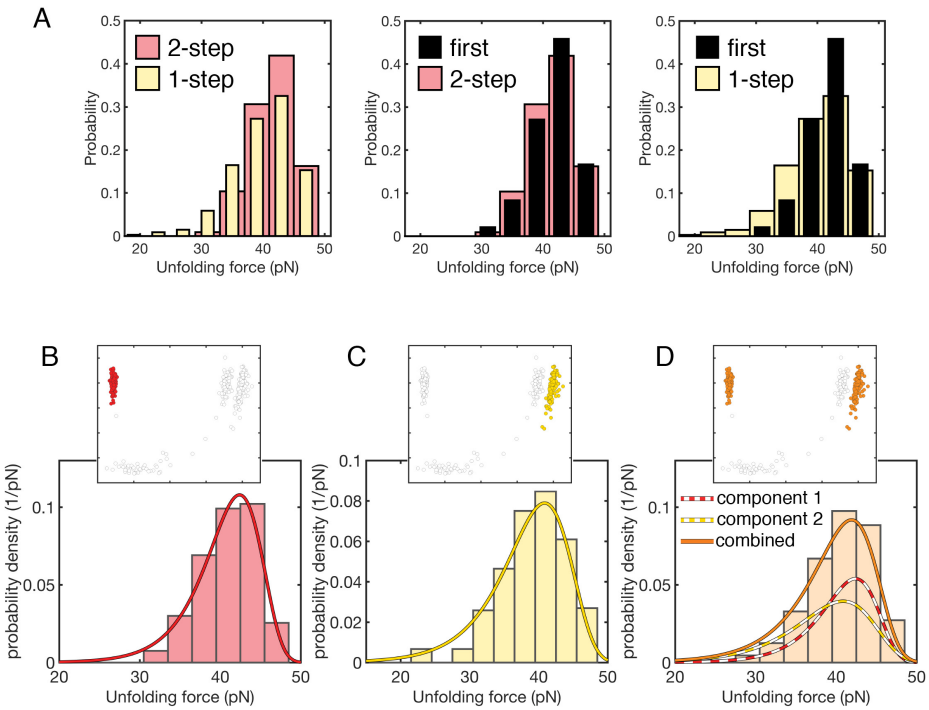


Figure IV-8 G-domain populates two distinct structured states. (A) Histograms of unfolding force distributions comparing two-step events (red bars) and one-step events (yellow bars) to the unfolding force measured during the initial stretching of the G-domain (black bars). For two-step events, only the unfolding force for the first step is considered. The distribution of the initial unfolding events is very similar to that of two-step events, but not to that of the one-step events. (B-D) Analysis of the unfolding force distributions for two-step events (B), one-step events (C), and combined events (D) from the 10 s refolding dataset. Red and yellow lines represent calculated probability densities using Bell's model, with parameters extracted by transforming the unfolding force distributions (red and yellow bars) into force-dependent lifetimes using the method of Dudko et al. (2008). The orange bars in D represent the entire unfolding force distribution and a calculated probability density (orange line) that is a combination of the two models in B and C. weighted by the probability of the individual components. The scatter plots on top highlight the data that was used for the respective analysis and plot on the bottom.

In order to resolve the two components observed during G-domain unfolding and characterize them individually, I first focused on only the 2-step transitions, which can easily be identified and represent component 1. Analysis of the force distribution for the first transition of 2-step events using Bell's model (Bell, 1978; Dudko et al., 2008) yielded a distance to the transition state for unfolding of $\Delta x^{\ddagger}_{unfolding}$ = 1.20 nm ± 0.08 nm and a native state lifetime of τ_0 = 113653 s ± 92012 s (Figure IV-8B). The large uncertainty for the value of the native state

lifetime is in part due to the relatively large unfolding forces, which requires long extrapolation to zero force.

I next estimated the parameters of the second unfolding component. I assume that the 1-step transitions represent a mixture of components 1 and 2 due to missed detection of the unfolding intermediate with a probability of $p_{miss} = 0.14$ (see above). I calculated the fraction of events that component 1 contributes to the 1-step transitions and corrected the unfolding force distribution of 1-step events accordingly (see Table IV-2). Analysis of the resulting unfolding force distribution

data set	$N_{1\text{-step}}$	$N_{2\text{-step}}$	f_{C2}	α
1 s refolding time	110	57	0.92	0.60
10 s refolding time	115	87	0.88	0.50
15 s refolding time	129	78	0.90	0.56

Table IV-2 G-domain two components analysis.

(Figure IV-8) yields parameters of $\Delta x^{\ddagger}_{unfolding} = 0.90$ nm ± 0.06 nm and $\tau_0 = 4956$ s ± 2606 s for the data set with a 10 s refolding pause. The fraction of component 2 in the complete data set from 10 s refolding time is $\alpha = 0.5$ (Table IV-2). Using these parameters, I am able to recapitulate the total distribution of unfolding forces reasonably well (Figure IV-8).

Thus, two components with similar transition state distances, but different zero force lifetimes are needed to describe unfolding of the G-domain. When we compare the three data sets with different refolding times (1 s, 10 s and 15 s), we find similar parameters for the component 2 transition state distance and lifetime (Supplementary Material, Table IV-3). The probability of component 2 also remained relatively constant over the variation of refolding time. we obtained $\alpha = 0.60$ for a refolding pause of 1 s and $\alpha = 0.56$ for a refolding pause of 15 s

(compared to $\alpha = 0.5$ for a 10 s refolding pause). We think these values are not significantly different therefore the component fraction remains constant with different folding time.

data set	$\Delta x^{\ddagger}_{unfolding}$ (nm)	τ_0 (s)
1 s refolding time	0.96 ± 0.06	8000 ± 4214
10 s refolding time	0.90 ± 0.06	4956 ± 2606
15 s refolding time	0.94 ± 0.04	7649 ± 2530
combined	0.94 ± 0.06	7539 ± 3768

Table IV-3 G-domain two components fitting results.

Our analysis indicates that component 2 is observed in ~55% of all unfolding events. At least two scenarios can explain the observation of these components: The natively folded G-domain could unfold through two different pathways, each with its own unfolding force distribution. Alternatively, the G-domain could adopt two distinct states during refolding, each with its own mechanical stability. To distinguish between these two possibilities, I carried out experiments in which I set the upper force limit to 30 pN. Due to the differences in unfolding force distributions described above, the probability of observing component 1 unfolding is lower than that of observing component 2 unfolding ($p_{unfolding}(C1,F\leq30\text{ pN}) = 0.03$ and $p_{unfolding}(C2,F\leq30\text{ pN}) = 0.07$). In the alternative pathways scenario, the lower force-route is available during each pull, and I would thus expect unfolding at the higher of the two frequencies. If alternative states account for the mixed distributions discussed above, unfolding in this force range should be rare, since the two components do not appear to quickly interconvert (see above results for different refolding times). When I repeatedly stretched the folded G-domain to 30 pN in force ramp experiments, I observed unfolding with a probability of 0.016 (5 events in 314 attempts, 5 molecules), much lower than what is expected for component

2 (p-value = 0.002 from Student's t-test) and for the entire set of unfolding data, which has an overall probability of unfolding below 30 pN of 0.04. This finding makes it unlikely that component 2 represents an alternative unfolding pathway.

Importantly, the force distribution of only the initial unfolding transition for each molecule (52 molecules total), which presumably occurs from the natively folded state, matches much better with the 2-step unfolding forces than the with the total unfolding force distribution (Figure IV-8A). It therefore appears that, after initial mechanical unfolding, the G-domain can refold into at least two states with indistinguishable length changes upon unfolding, but different mechanical stabilities. The fact that initial unfolding occurs with a force distribution similar to component 1 makes it likely that component 1 represents the native state, whereas component 2 represents a non-native state that is adopted in about half of all refolding attempts.

IV-2.6: The ribosome slows G-domain folding

The autonomous folding of the G-domain makes it likely that folding begins as soon as the domain emerges from the ribosome during EF-G synthesis. The ribosome has previously been shown to modulate nascent protein folding (Kaiser et al., 2011). We therefore asked whether the ribosome also has an effect on the nascent G-domain. To this end, I generated ribosome-nascent chain complexes (Figure IV-9A) using an in vitro translation system that was programmed with in vitro transcribed mRNA (Kaiser et al., 2011). The mRNA terminated at codon 328 of the EF-G sequence (CCG, encoding proline). Due to the absence of a stop codon, the translation product accumulates as a peptidyl-tRNA (Figure IV-9B) that is stably anchored in the ribosomal P-site. With the arrest position 35 amino acids downstream of the G-domain boundary, domain II serves as a natural spacer that allows the entire G-domain to emerge from

the ribosome (see Figure IV-9A). I refer to this construct as the G-domain ribosome-nascent chain complex (G-RNC/328_{RNC}).

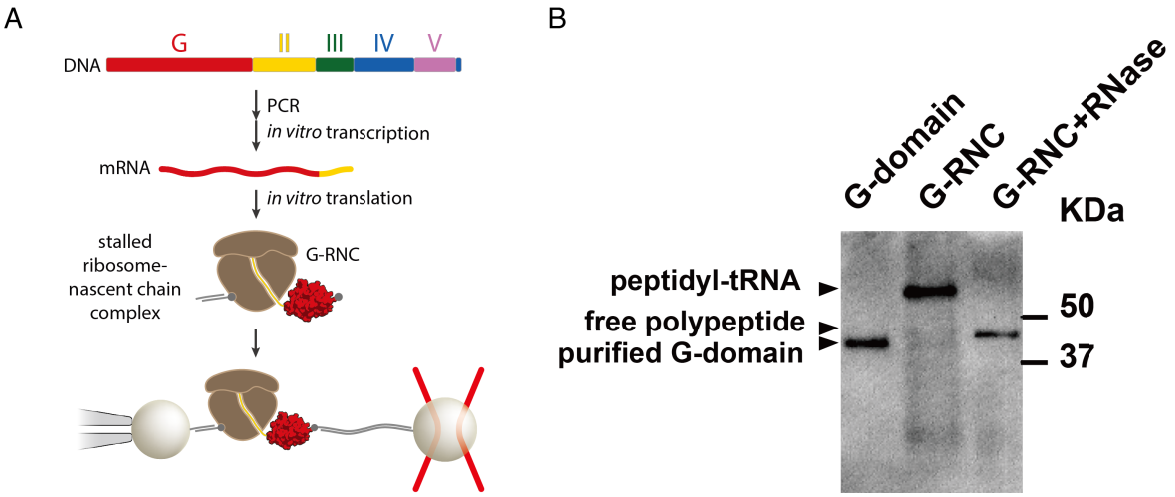


Figure IV-9 Stalled 328_{RNC} sample preparation. (A) Schematic of the experimental procedure for analysis of 328_{RNC}. Stably stalled ribosome-nascent chain complexes are produced by *in vitro* transcription and translation. Isolated complexes are modified to create attachment points and tethered for optical tweezers experiments. (B) Western blot analysis of in vitro translated G-RNC (328_{RNC}). The isolated G-domain and 328_{RNC} samples were separated on an SDS-PAGE gel. Proteins were electro-blotted onto a nitrocellulose membrane and detected with streptavidin-HRP. Lane 1: purified G-domain (serving as control), lane 2: G-RNC, lane 3: G-RNC treated with RNase A. The untreated ribosome-bound nascent chain as a much lower electrophoretic mobility than the free protein. RNase treatment results in a shift to higher electrophoretic mobility due to removal of the tRNA moiety. This result confirms that the translation product accumulates as peptidyl-tRNA, as expected for stalled RNCs.

The transitions observed upon G-RNC unfolding (Figure IV-10) are similar to those of the isolated G-domain: They mostly occurred between 30 and 50 pN with a total contour length change of 103.3 nm ± 1.7 nm (std), often composed of individual transitions contributing 11.2 nm ± 1.7 nm and 92.2 ± 2.0 nm. The observation that G-RNC populates the same unfolding intermediate detected in the isolated native G-domain suggests that the protein adopts the same folded structures on the ribosome and in isolation. In addition to the mechanically labile states that unfold below 10 pN, G-RNC populates structures that unfold in the 10 to 30 pN range (Figure IV-10). Several of these events are correlated (i.e., they occur in close succession in the same trace, see Figure IV-10), with combined contour length changes often similar to unfolding

of the full G-domain. These states are not observed with the isolated G-domain and could represent folding intermediates that rapidly progress to the fully folded state in the absence of the ribosome, but accumulate in G-RNC due to a deceleration of the forward folding rate.

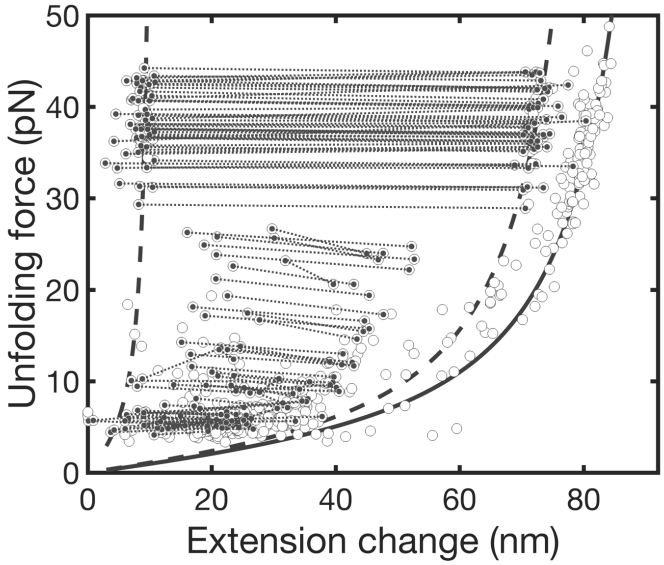


Figure IV-10 Force-extension change scatter plot for G-RNC unfolding. All transitions obtained when repeatedly stretching G-RNC are plotted as open circles. The 2-step transitions are further plotted as solid small circles, with connected lines indicating that they are sequential transitions from the same stretching curve. The native 2-step unfolding is well separated from the non-native 2-step transitions in force as well as the extension changes.

The overall folding toward the stable structures is markedly slower for G-RNC than it is for the G-domain alone. Under identical refolding conditions, G-RNC folds into stable structures much less frequently than the isolated G-domain. I observed native state unfolding with an overall probability of p_{refold} (G-RNC, 10 s, 2 pN) = 0.27 after a refolding pause of 10 seconds at 2 pN, a value similar to that obtained for the isolated G-domain after a 1 second pause (p_{refold}(G-domain, 1 s, 2 pN) = 0.32). Thus, overall folding of G-RNC is decelerated approximately 10-fold compared to the isolated protein.

IV-2.7: The ribosome destabilizes the native and the non-native structures

In addition to slower folding, G-RNC also exhibits unfolding at lower forces than the isolated G-domain (Figure IV-11A), suggesting that the ribosome mechanically destabilizes the natively folded structure. This destabilization is consistent with a lower value for the zero-force lifetime of the natively folded ribosome-bound domain from analyzing the unfolding-force distribution of 2-step events (Figure IV-11B). I obtain a value of τ_0 = 78300 s ± 10815 s, while the distance to the transition state for

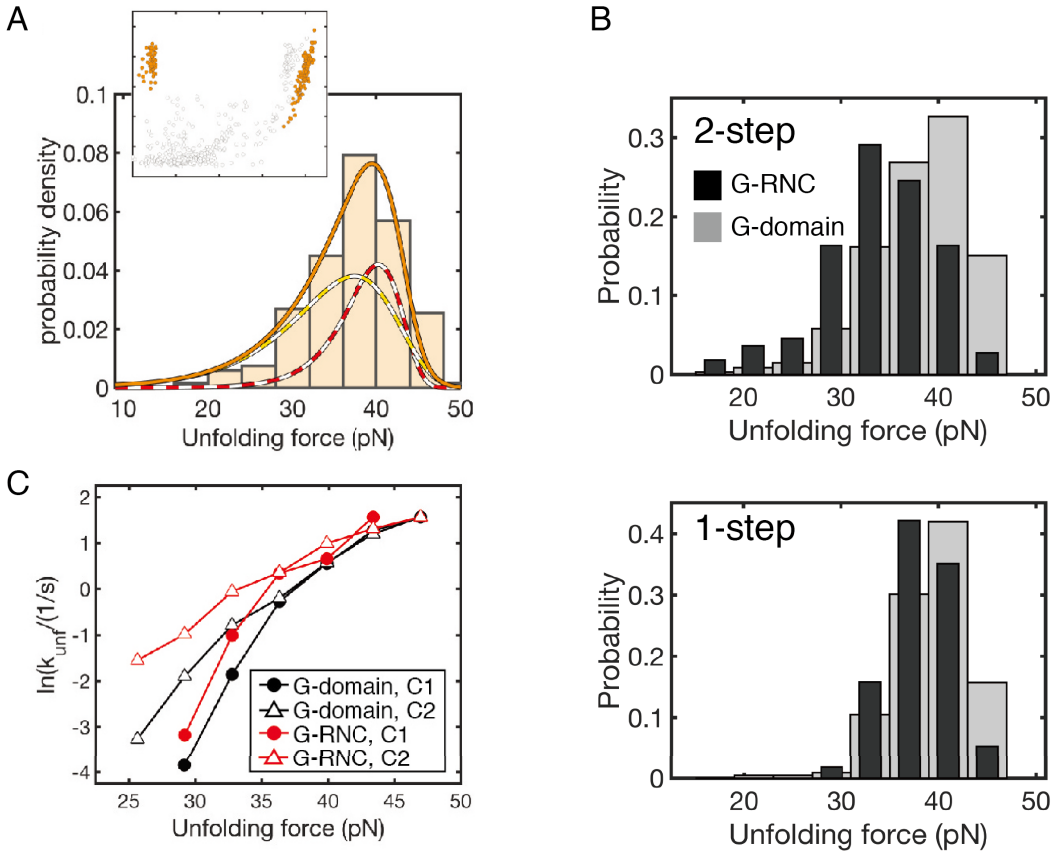


Figure IV-11 The ribosome destabilizes folded structures of nascent G-domain. (A) Unfolding force distribution and model for the two components that characterize unfolding as in Figure IV-8. (B) Comparison of the unfolding force distribution of G-RNC (black bars) and the isolated G-domain (grey bars). G-RNC exhibits lower unfolding forces for both 2-step transitions (top) and 1-step transitions (bottom), indicating that the native and the non-native states are destabilized by the ribosome. (C) Comparison of the force dependent unfolding rates (k_{unf}) for G-RNC (red) and the G-domain (black), separated by components (C_1 and C_2). The ribosome destabilizes both

components to similar extents. The ln(k_{unf}) vs F plots exhibit rollover at high force, indicating that the barrier for unfolding begins to disappear in this regime.

unfolding, $\Delta x^{\ddagger}_{unfolding}$ = 1.23 nm ± 0.15 nm, is similar for both the G-domain and G-RNC. These results indicate that the G-RNC has a slightly higher unfolding rate than the G-domain alone, at least in the experimentally accessible force range.

Similar to what we observed for isolated protein, G-RNC appears to populate two stable states. The force distribution for 1-step unfolding events is well described by a two-component model (Figure IV-11A) as described above for the free G-domain. The second component has parameters of $\Delta x^{\ddagger}_{unfolding}$ = 0.68 nm ± 0.05 nm and τ_0 = 388 s ± 152 s. Thus, the second component has a zero-force lifetime much shorter than the one determined for the G-domain alone, suggesting that the ribosome destabilizes the non-native state. The probability of component 2 in the distribution is α = 0.62, similar to what we observed for the isolated G-domain, indicating that the presence of the ribosome does not alter the probability of folding into either the native state (component 1) or the non-native state (component 2). Consistent with this observation is the fraction of 2-step events that are unique to the native state, $f_{2\text{-}step}$ = 0.33, similar to the value determined in the absence of the ribosome (see above). In summary, the ribosome affects several aspects of G-domain folding. It destabilizes both the native and the non-native state and reduces the overall forward folding rates.

IV-3: Discussion

Here, I have made a first step toward developing elongation factor G, EF-G, as a model to study multidomain protein folding and the influence of the ribosome on the process, using mechanical manipulation of individual molecules with optical tweezers. Since EF-G folding has not previously been studied, I first characterized the properties of individual domains in the

context of the full-length protein, and then proceeded to characterize how folding of the N-terminal G-domain, which may represent a crucial step during the biogenesis of the full-length protein, is affected by the ribosome.

Unfolding of full-length EF-G proceeds from the C-terminal portion of the molecule and occurs mostly in a domain-wise fashion. The three C-terminal domains unravel at relatively low forces, well below 20 pN. While I cannot unambiguously assign these transitions to the individual domains, domain III appears to be mechanically weakest, unfolding below 15 pN. Domains IV and V are slight more stable. They often unfold in one apparent step, indicating energetic coupling. In fact, domain IV is discontinuous in sequence, straddling domain V. This arrangement is reminiscent of the subdomain arrangement in T4 lysozyme, where a similar topological coupling results in folding cooperativity important for efficient folding (Shank et al., 2010).

The two N-terminal domains, domain II and the G-domain, unfold at high forces. Most likely, the extensive interface between the two domains ruptures first, releasing amino acids 294 to 317 in the loosely structured region just upstream of the first strand of the domain II beta sheet. Based on the crystal structure, such a rupture would result in a contour length change of 8.9 nm, very close to the observed first transition during domain II unfolding (9 nm). The resulting partially structured domain II is highly unstable and quickly unfolds (chapter V). Thus, the G-domain and domain II are energetically coupled, although through a mechanism that is different from the one described for domains IV and V above.

Attempts to refold the full-length protein did not result in complete renaturation. In a small fraction of attempts, domains III, IV, and V appeared to regain their initial structure. Even

more rarely, the G-domain refolded. Overall, folding of the full-length protein was very inefficient, even at the level of individual domains. Inefficient folding of the G-domain in the full-length protein is in contrast to robust refolding of the same domain in isolation.

Perhaps several unfolded domains form non-native, off-pathway structures that need to unfold for productive folding to proceed. The observation of transitions that are longer than any individual domain (Figure IV-5) supports this interpretation. Similar effects have been observed for heat shock protein 90 (Jahn et al., 2016), another multidomain protein. During folding in vivo, domain-wise folding during synthesis likely avoids a scenario in which all domains in a given polypeptide molecule are simultaneously unfolded. Notably, I did not observe a single occurrence of domain II refolding. It is possible that domain II requires contacts with the structured G-domain to be stably folded, given the extensive interface between the two domains in the native structure. A similar scenario has been observed for phosphoglycerate kinase, whose C-terminal domain requires the N-terminal domain to fold (Young et al., 2007), perhaps reflecting a general hierarchy of domain-wise folding from N- to C-terminus. If properly formed upstream structures are important for subsequent folding, it would be particularly important that they complete their folding in an efficient and timely manner to avoid accumulation of unfolded domains that enter into detrimental interactions.

The N-terminal G-domain indeed folds autonomously as soon as it emerges from the ribosome during EF-G synthesis. When the G-domain folds in close proximity to the ribosome, overall refolding rates are reduced approximately 10-fold. The ratio of the native to the non-native state does not appear to be affected, suggesting that the pathways leading to these states branch after common initial steps, and that the ribosome affects steps before branching. Reduced

folding rates may generally help to delay folding of individual domains or small proteins during synthesis until sufficient sequence has been produced for the nascent polypeptide to enter into productive folding pathways. In the context of multi-domain proteins, an additional function could be the stabilization of partially structured states competent to form native interfaces with other domains. In the related scenario of forming a complex between structures encoded on separate polypeptides, the ribosome - together with molecular chaperones - has been shown to play a crucial role (Fedorov and Baldwin, 1995; Shieh et al., 2015). Folding of domain II and the formation of the G-domain-domain II structural unit might require similar contributions, which would explain why it fails to fold during renaturation of full-length EF-G.

In addition to decreasing forward folding rates, the ribosome also decreases the (mechanical) stabilities of the native and non-native G-domain states, suggesting increased unfolding rates. Destabilization of structured nascent proteins by the ribosome has previously been observed by us (Kaiser et al., 2011) and others (Cabrita et al., 2016; Samelson et al., 2016). Our optical tweezers experiments are non-equilibrium measurements that do not allow us to directly determine thermodynamic stabilities. The observation that the ribosome decreases G-domain folding rates while at the same time increasing the rate of unfolding could suggest a thermodynamic destabilization, which would be in agreement with direct thermodynamic measurements of nascent chain stabilities by NMR spectroscopy (Cabrita et al., 2016) and proteolysis (Samelson et al., 2016). Structural destabilization is thus emerging as a general effect of the ribosome on nascent polypeptides, at least in the case of natural proteins. The folding of the artificially designed Top7 protein (Kuhlman et al., 2003), on the other hand, is not detectably affected by the ribosome (Goldman et al., 2015), raising the possibility that destabilization by the

ribosome is an evolved property of natural proteins with significance for de novo folding. However, few proteins have been investigated in this respect so far, and a much larger data set is required to evaluate this hypothesis.

Interestingly, our unfolding experiments reveal a dichotomy within the G-domain unfolding transitions. Multiple parallel unfolding pathways could in principle explain this behavior and have in fact been observed experimentally for the src SH3 domain (Zhuravlev et al., 2016). However, we do not observe upward curvature in a plot of $\ln(k_{unf})$ vs. force (Figure IV-11C), a hallmark of this behavior. Instead, our results indicate that the G-domain populates two distinct states. The natively folded state is characterized by a narrower force distribution centered around a higher mean force and often populates a short-lived unfolding intermediate. The non-native state unfolds at lower force without populating this intermediate and is either not yet fully structured or has entered a misfolded state. On the time scale we were able to follow refolding (1 to 15 s), the relative population of the non-native state does not decrease, even though the fraction of structured states (both native and non-native) increases during this time interval. Thus, the non-native state likely represents a misfolded or trapped state that does not directly convert into the native structure, at least not on the time scale of overall productive folding. Interestingly, the non-native state is destabilized more than the native state, as judged by the shift in the unfolding force distributions. Thus, the ribosome does not prevent the formation of this state, but facilitates its unfolding. In the cell, ribosome-binding molecular chaperones (Kim et al., 2013) like trigger factor probably further destabilize misfolded states. Contributions from the ribosome may enhance this important biological activity.

The structural identity of the non-native G-domain state is not clear. It unfolds with an extension change indistinguishable from that of the native state at only slightly lower forces. This observation could suggest that the termini of the protein are in a similar configuration in both states, but form a different network of interactions with the rest of the domain. We note that our construct (residues 1-294 of EF-G) contains 11 proline residues, which are all in the trans conformation in the EF-G crystal structure. It is possible that the non-native G-domain structure that we observe in our experiments contains cis prolines. Destabilization of the native structure by the ribosome could help to avoid the formation of stable structure containing incorrect proline conformers, facilitating isomerization to the correct conformers by peptidyl-prolyl isomerases such as the ribosome binding-chaperone trigger factor. Additional studies are required to unambiguously determine the structural identity of the non-native state and its significance for overall productive folding.

Taken together, our studies demonstrate that the previously observed destabilization of nascent polypeptide structures occurs in an authentic globular multidomain protein, EF-G. This contribution of the ribosome may help efficient folding of the G-domain, the first domain to emerge during EF-G synthesis, whose folding might represent an important waypoint en route to the native full-length protein.

IV-4: Materials and methods

III-4.1: Transforming unfolding force distribution to force-dependent rates

To extract the lifetime and transition state distance at zero force for the G-domain and G-RNC, I applied the method developed by Dudko *et al.* (Dudko et al., 2008). Force dependent loading rates were either directly determined from our experimental data or modeled using the

equation described in (Dudko et al., 2008). I then fitted our data using Bell's model (Bell, 1978) to extract the transition state distance and zero force lifetime.

III-4.2: Mixed-component analysis

To calculate the fraction of component 2 among 1-step events and correction of the unfolding force probability density, the fraction of events from component 2 in the 1-step unfolding transitions, f_{c2}, can be expressed in terms of the total number of 1-step transitions ($N_{1\text{-}step}$) and 2-step transitions ($N_{2\text{-}step}$) and the probability of missing detection of the unfolding intermediate (p_{miss}) that results from the intermediate lifetime being of a similar order of magnitude as our sampling rate. The number of 1-strep unfolding events is $N_{1\text{-}step} = N_{c2} + p_{miss} * N_{2\text{-}step}/(1 - p_{miss})$. The fraction of component 2 events in the 1-step transitions is $f_{c2} = N_{c2}/N_{1\text{-}step} = 1 - (p_{miss} * N_{2\text{-}step})/((1 - p_{miss}) * N_{1\text{-}step})$. Assuming that the 2-step events are purely from component 1, the observed probability density of unfolding forces for the 1-step transitions, $p_{1\text{-}step}$, can then be expressed as a combination of probability densities of the two components, $p_{1\text{-}step} = f_{c2} * p_{c2} + (1\text{-}f_{c2}) * p_{c1}$, where p_{c1} is the probability density of component 1. Therefore, the probability density of component 2 $p_{c2} = 1/f_{c2} * (p_{1\text{-}step} - (1 - f_{c2}) * p_{c1})$. Using this relationship, we numerically correlated the experimentally determined probability density of 1-step unfolding forces for the small contribution of component 2.

To calculate the fraction of component 2 in the complete unfolding force distribution, an $\alpha = N_{c2}/N_{total}$ was defined as the fraction of the number of transitions from component 2, N_{c2}, in total number of transitions, $N_{totoal} = N_{1\text{-}step} + N_{2\text{-}step} = N_{c1} + N_{c2}$. Assuming that all 2-step transitions represent component 1, and with the totoal number of component 1 transitions given by $N_{c1} = N_{2\text{-}step}/(1 - p_{miss})$, the fraction of component 2 in the distribution was calculated as $\alpha = 1$

– $N_{2\text{-step}}/(N_{total} * (1 - p_{miss}))$. For the complete data set including all refolding times, the weighted average is $\alpha = 0.55$, therefore 55% of all unfolding events are contributed by component 2.

To extract unfolding parameters from the unfolding force distribution, I analyzed the unfolding force distributions by converting them into force-dependent lifetimes using calculated loading rates (trap-stiffness k = 0.09 pN/nm, handle stiffness for 1900 bp dsDNA handle calculated using the worm-like chain model with a persistence length of p = 53 and L_c = 0.34 nm/bp).

Chapter V: The ribosome cooperates with a chaperone to guide multidomain protein folding

By

Christian Kaiser and Kaixian Liu

This work has been submitted to Molecular Cell as:

Liu, K., Maciuba, Kevin, Kaiser, C.M. (2019). The ribosome cooperates with a chaperone to guide multi-domain protein folding. *Molecular Cell*, 74, 310-319

Contributions: Kaixian Liu conducted most of the experiments. Kevin Maciuba performed most of the 358-, 386- and 415-RNCs force ramp experiments, some of the dentruation experiments and some force ramp experiments with TF. All authors analyzed the data. Kaixian Liu and Christian Kaiser created figures, and wrote the manuscript.

Chapter V. The ribosome cooperates with a chaperone to guide multidomain protein folding

V-1: Introduction

Combining several functional domains into a single polypeptide is a widespread evolutionary strategy for generating proteins with novel functions (Han et al., 2007). How these complex proteins fold into the specific three-dimensional structures that are required for function remains poorly understood. Beginning with the pioneering experiments by Anfinsen (Anfinsen, 1973), protein folding mechanisms have chiefly been studied using small proteins that fold rapidly and reversibly (Braselmann et al., 2013). Even though many of the mechanistic details remain elusive, several general principles have emerged from these studies (Dill and MacCallum, 2012; Sosnick and Barrick, 2011). For instance, folding transitions are largely cooperative, but "molten globule" intermediates, containing secondary structure while lacking stable tertiary interactions are frequently populated along folding pathways. Comparatively little is known about the folding mechanisms of multidomain proteins (Jahn et al., 2016; Walters et al., 2013). The population of multiple intermediate states (Brockwell and Radford, 2007) and slow overall folding result in a high propensity of multidomain proteins to form insoluble aggregates *in vitro*, hampering mechanistic folding studies.

Folding of multidomain proteins begins cotranslationally (Frydman et al., 1999; Nicola et al., 1999), while the ribosome still elongates the nascent polypeptide chain. As a consequence, these proteins acquire stable structure during synthesis, and cotranslational folding has long been recognized as a potential mechanism of facilitating their conformational search (Fedorov and Baldwin, 1997). However, it remains unclear to what degree sequential domain-wise folding is

necessary for productive folding of natural multidomain proteins (Han et al., 2007), whether a particular folding order is required (Batey and Clarke, 2008), and how well biophysical studies of small single-domain proteins reflect folding events in larger multidomain proteins (Braselmann et al., 2013). In addition, it has recently become apparent that interactions with the ribosome decrease the folding rate (Kaiser et al., 2011) and the stability (Cabrita et al., 2016; Holtkamp et al., 2015; Samelson et al., 2016) of cotranslationally formed structures. Whether and how these interactions contribute to productive folding remains unknown.

Molecular chaperones are crucial for cellular protein folding and maintenance, preventing aggregation and guiding the conformational search (Kim et al., 2013). Specialized nascent chain-binding chaperones interact with polypeptides as soon as they emerge from the ribosome during synthesis (Kramer et al., 2009). In bacteria, the chaperone trigger factor (TF) binds to the ribosome near the polypeptide exit tunnel (Kramer et al., 2002). As such, TF is the first chaperone to interact with emerging nascent proteins and, together with the bacterial Hsp70 system, helps the folding of newly synthesized proteins (Genevaux et al., 2004). TF has been reported to increase folding yield by reducing folding rates (Agashe et al., 2004; Hoffmann et al., 2012). However, a mechanistic understanding of TF action is lacking. More generally, the importance of several chaperones for the biogenesis of functional proteins is firmly established (Kim et al., 2013), but it remains largely unclear how they contribute to productive folding.

Experimentally studying early folding of nascent multidomain proteins is challenging. Many of the optical methods that are traditionally used to follow protein folding, such as circular dichroism spectroscopy or tryptophan fluorescence measurements, cannot be applied in the presence of the ribosome and molecular chaperones. Moreover, transiently populated and

potentially heterogeneous states along the folding pathway often remain unresolved in ensemble measurements. Mechanical manipulation of individual protein molecules with optical tweezers is a powerful approach for following folding directly (Zoldak and Rief, 2013), yielding folding rates and one-dimensional structural information. This single-molecule technique is ideally suited for studies of complex proteins (Jahn et al., 2016; Liu et al., 2017; Stigler et al., 2011; Yu et al., 2012) and circumvents the complication of protein aggregation.

I have used optical tweezers to explore early folding steps during the synthesis of elongation factor G (EF-G), a GTPase that is required for efficient polypeptide elongation. EF-G is composed of five domains (Czworkowski et al., 1994) and has orthologs in all kingdoms of life. I have previously reported (Liu et al., 2017) that interactions among unfolded domains globally interfere with productive folding of EF-G. Here, I show that folding of the N-terminal G-domain of EF-G has to precede folding of the following domain II, imposing a hierarchical folding order. Depending on nascent chain length, the ribosome either accelerates or decelerates G-domain folding. Misfolding between the two domains slows productive folding, but is ameliorated by the ribosome and the chaperone TF. Surprisingly, I found that natively folded structures can be denatured by interactions with unfolded parts of the nascent protein, resulting in an unanticipated complication of folding. TF, but not the ribosome, prevents this denaturation. Our studies showed how the ribosome and the chaperone together guide early folding to set the protein on the right path for productive folding.

V-2: Results

V-2.1: The ribosome modulates G-domain folding

Domain-wise folding during protein synthesis represents a conceptually straightforward way of simplifying the conformational search for the native state. Stalled ribosome-nascent chain complexes (RNCs) of EF-G were generated to study early cotranslational folding with optical tweezers (Figure. IV-9A). When translation is stalled at codon 328 of the EF-G coding sequence (328_{RNC}), the N-terminal G-domain (aa 1-293 of EF-G) has fully emerged from the ribosome, whereas the following 35 residues (aa 294-328) of domain II span the exit tunnel in the large ribosomal subunit (Figure V-1). 328_{RNC} thus represents a "snapshot" of EF-G synthesis where the first domain has emerged from the ribosome and can fully fold.

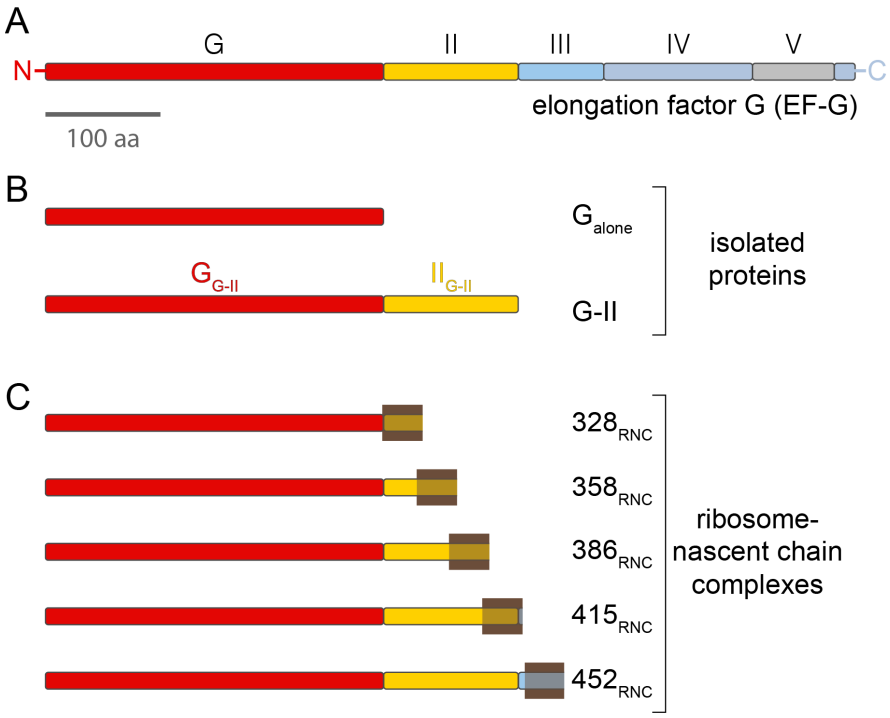


Figure V-1 EF-G constructs primary structure diagram. (A) Full length EF-G primary structure diagram with length bar indicated. Domain names indicated on top. (B) Diagrams of the two EF-G constructs studied as isolated polypeptides in the absence of the ribosome. (C) Diagrams of nascent chains. The brown block indicates the approximate position of the ribosome exit tunnel, i.e. the parts of the nascent EF-G polypeptides inside the box are

sequestered within the exit tunnel (assuming that 35 amino acids are contained within the tunnel), the parts to the left of the boxes have fully emerged from the ribosome. In (B) and (C), the nomenclature for the respective construct is indicated on the right.

I applied a continuously increasing force ("force ramp") to individual *328*$_{RNC}$ molecules. The resulting force extension curves (FECs) showed transitions that indicated cooperative unfolding of the natively structured G-domain (Fig. 1B, red trace). Measuring the unfolding force and extension change (Figure V-2) of the initial unfolding transitions, I obtained a contour length change of ΔL = 103.2 nm ± 1.8 nm (mean ± standard deviation, std; 30 molecules). This value is very close to the

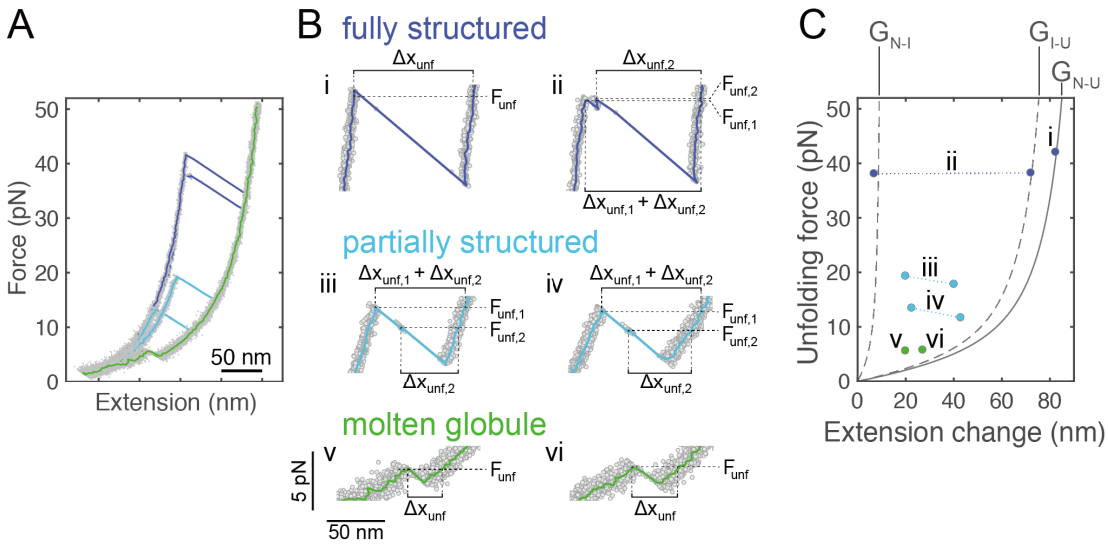


Figure V-2 Example transitions of *328*$_{RNC}$. (A) Overlay of six FECs recorded with a single *328*$_{RNC}$ molecule. Grey dots: raw data (1000 Hz), colored lines: time-averaged data (33 Hz). The data was chosen to represent two examples each of the molten globule-like (green), partially structured (cyan) and fully structured (blue) states. (B) Magnified view of the unfolding transitions for the individual traces shown in A. Each transition is characterized by the unfolding force (F_{unf}) and extension change (Δx_{unf}). Unfolding occurs either directly in one step (panels I, v, vi) or via an intermediate in two steps (ii, iii, iv). The number of unfolding steps, the unfolding force range and the extension change are characteristics of the protein structure that unfolds. (C) Scatter plot of unfolding transitions. Each transition is represented by a circle. Transitions from two-step unfolding, resulting from the population of an unfolding intermediate, are connected by dotted lines. This representation enables visual comparison of a large number of unfolding events. Grey lines represent worm-like chain (WLC) models (Bustamante et al., 1994) of one-step unfolding (solid line) and two-step unfolding (dashed lines)., which are characteristic for the fully folded G-domain of EF-G (Liu et al., 2017). A persistence length of 0.65 nm and contour length of 0.36 nm per amino acid residue were used for polypeptide WLC models.

expected contour length change of ΔL_{calc} = 103.7 nm, which is consistent with G_{alone}, indicating that the G-domain adopts its native structure in 328_{RNC} and suggesting that its folding is a waypoint toward native EF-G.

After complete unfolding, I relaxed the force to 2 pN for 10 seconds to permit refolding. Subsequently, I pulled again to probe which structures had formed. Repeated cycles of pulling and relaxation yielded heterogeneous transitions (Figure V-2). The G-domain refolded in 35.8% of attempts, as indicated by unfolding transitions similar to the initial one (Figure V-2). In addition, mechanically weak molten-globule-like structures (Cecconi et al., 2005) and partially folded intermediates are observed frequently in 328_{RNC} (Figure V-3). These states are detected much more rarely in an isolated polypeptide comprising the G-domain (termed G_{alone} here) (Figure V-3). Instead, G_{alone} exhibits a higher fraction of fully folded structures, suggesting that folding proceeds more rapidly in the isolated protein than on the ribosome.

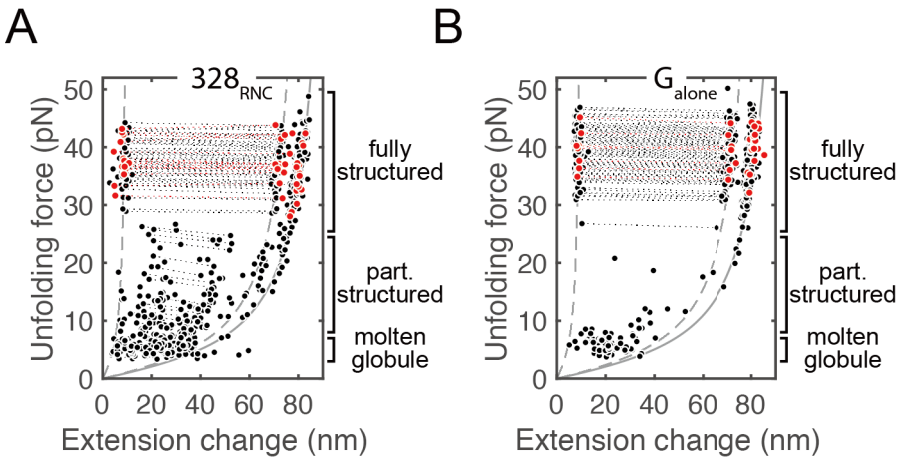


Figure V-3 Force-extension change plots comparison between 328_{RNC} and G_{alone}. (A) scatter plot of unfolding transitions for 328_{RNC}. Red dots: initial pull, black dots: after refolding. Dotted lines connect transitions from two-step unfolding. Only ~36% of traces after refolding how transition similar to the initial ones (labeled "fully structured"), indicating that the G-domain folds to the native structure in only about one third of attempts during the refolding 10-second refolding pause. Transitions that do not match this pattern represent partially structured and molten globule-like state ("part. Structured", "molten globule"). (B) Scatter plot of unfolding transitions of G_{alone}. Most transitions fall into the "fully structured" regions, indicating that G_{alone} reaches the native state with high

probability during the 10 second refolding pause. Colors and labeling as in panel A. Dashed and solid grey lines represent WLC models for two-step and one-step unfolding.

The population of several partially structured states (Figure V-3) suggests an intricate folding mechanism for the G-domain that is challenging to fully characterize. I quantified overall folding rates by analyzing the formation of fully structured states during repeated force ramp cycles (Figure V-4). By fitting the cumulative folding probability to a geometric distribution (Figure V-4C), I determined an apparent folding rate of $k_{app}(328_{RNC}) = 0.044\ s^{-1}$ (95% confidence intervals, ci: $0.040\ s^{-1}$, $0.048\ s^{-1}$) for 328_{RNC}, slower than that of the free protein ($k_{app}(G_{alone}) = 0.149\ s^{-1}$ (ci: $0.136\ s^{-1}$, $0.165\ s^{-1}$)) (Figure V-5). Thus, the overall G-domain folding rate is reduced in close proximity to the ribosome, presumably due to interactions of the ribosome with the nascent polypeptide. Such a deceleration of folding is consistent with previous reports (Kaiser et al., 2011; Liu et al., 2017).

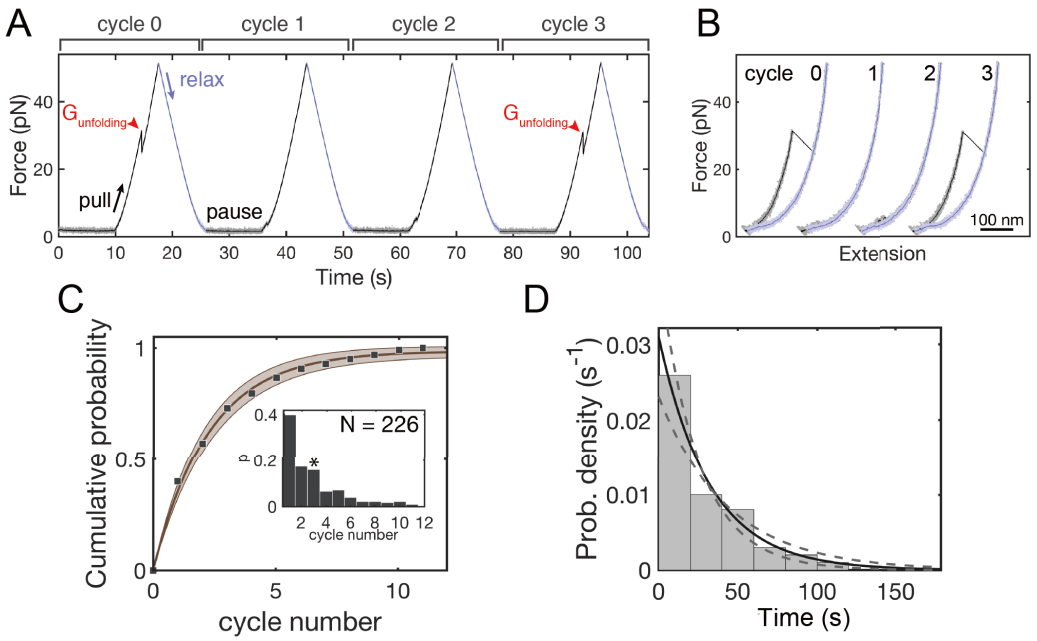


Figure V-4 Folding cycle analysis. (A) Example plot of force vs. time for a 328_{RNC} molecule subjected to force ramp cycles. We define cycles as beginning at the 2 pN refolding pause and ending after the subsequent relaxation. Unfolding of the structured G-domain (red arrowheads) is apparent in the initial cycle and in cycle 3. Stretching is shown in black, relaxation in blue. (B) The same cycles that are shown in A, represented as FECs. Cycles are offset

along the x-axis for clarity. Arrowheads indicate the unfolding events of the fully structured G-domain that are also shown in A. (C) Fit of geometric distribution to G-domain folding probabilities for *328*$_{RNC}$. Folding probability (black squares) is obtained from folding cycle analysis. The axis represents the cycle number in which unfolding occurs. Solid lines and shaded areas represent the mean and 95% confidence intervals of the geometric distribution fit to the data. The inset is the distribution of the folding cycle (raw counts) for *328*$_{RNC}$. The asterisk indicates the bin to which example in panel A contributed.

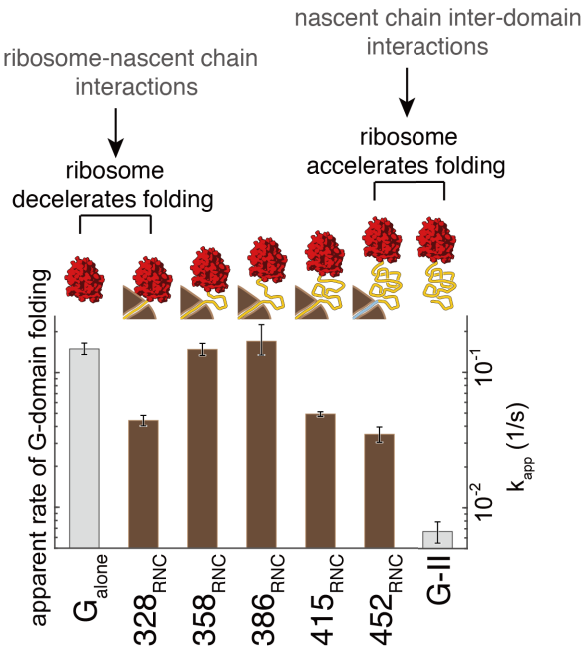


Figure V-5 Folding rates summary for all EF-G constructs. Apparent folding rates of the G-domain for the isolated *G*$_{alone}$ polypeptide (left, grey bar) and RNCs with varying lengths (brown bars). The bar diagram and cartoons on top indicate the stalling positions for the respective constructs. The ribosome decelerates folding at short nascent chain lengths (*328*$_{RNC}$ vs *G*$_{alone}$). After reaching a maximum, folding rates drops in longer nascent chains dur to non-productive interactions within the nascent polypeptide. This interference is even more pronounced in the isolated *G-II* protein comprised of the G-domain and domain II (right, grey bar), indicating that the ribosome reduces non-productive inter-domain interactions.

Shifting the stalling position from codon 328 to codon 358, and further to codon 386 in the EF-G coding sequence (*358*$_{RNC}$ and *386*$_{RNC}$, respectively) gradually restored the G-domain folding rate [k_{app}(*358*$_{RNC}$) = 0.148 s^{-1} (ci: 0.134 s^{-1}, 0.164 s^{-1}; k_{app}(*386*$_{RNC}$) = 0.170 s^{-1} (ci: 0.135 s^{-1}, 0.225 s^{-1})] (Figure V-5). The polypeptide exit tunnel that traverses the large ribosomal subunit is approximately 100 Å long, accommodating about 35 amino acids of unfolded polypeptide. Therefore, approximately 60 aa of domain II have fully emerged from ribosome in *386*$_{RNC}$, and the effect of the ribosome on G-domain folding appears to be minimal. These findings suggest

that the distance-dependent modulation of folding kinetics observed previously with artificial T4 lysozyme nascent chains (Kaiser et al., 2011) is relevant during the folding of an authentic multidomain protein. During EF-G synthesis, folding of the N-terminal G-domain is hence expected to occur with high probability soon, although not immediately, after it has emerged from the ribosome.

V-2.2: Interference among domains slows down folding

Elongating the nascent chain further to include 415 and 452 residues of EF-G (*415_{RNC}* and *452_{RNC}*; Figure V-1) resulted in progressively slower folding rates $k_{app}(415_{RNC}) = 0.049\ s^{-1}$ (ci: $0.047\ s^{-1}$, $0.051\ s^{-1}$); $k_{app}(452_{RNC}) = 0.035\ s^{-1}$ (ci: $0.031\ s^{-1}$, $0.040\ s^{-1}$)] (Figure V-5). Since the ribosome does not modulate folding over long distances, interactions within the nascent polypeptide must be responsible for the observed effect. Most likely, the unfolded domains form off-pathway structures that slow down overall folding. Frustration of G-domain folding was even more pronounced in an isolated polypeptide containing the first two EF-G domains (aa 1-410 of EF-G, termed *G-II* here). Compared to *452_{RNC}*, the rate of G-domain folding in *G-II* is much reduced ($k_{app}(G_{G\text{-}II}) = 0.006\ s^{-1}$ (ci: $0.004\ s^{-1}$, $0.007\ s^{-1}$), Figure V-5).

These results imply that the G-domain is likely to fold within a narrow window during EF-G synthesis, with folding occurring with the highest probability when around 100 residues of domain II have been synthesized (*386_{RNC}*). Folding is slowed down by the ribosome in shorter nascent chains (*358_{RNC}*), and by interactions within the nascent polypeptide in longer ones (*415_{RNC}*, *452_{RNC}*). Previously, the ribosome was thought to exclusively delay folding (Cabrita et al., 2016; Kaiser et al., 2011). Here, it appears that the effect of the ribosome on the folding of a particular domain is context specific: when close to the ribosome, G-domain folding is slowed

(*328*$_{RNC}$ vs. G_{alone}). In longer nascent chains, the ribosome accelerates G-domain folding, relative to an analogous isolated polypeptide (*452*$_{RNC}$ vs. *G-II*). Therefore, presumably similar interactions between the ribosome and the nascent polypeptide result in distinct outcomes, depending on the stage of synthesis.

V-2.3: The ribosome and TF reduce inter-domain misfolding

To characterize the structures that the two-domain EF-G constructs populate during refolding in the presence (*452*$_{RNC}$) and absence (*G-II)* of the ribosome, I carried out "force clamp" experiments. I subjected these molecules alternating to high force (30 pN), favoring unfolding, and low force (3.5 pN), permitting refolding. When held at this low force, the isolated *G-II* polypeptide transitions between several states (Figure V-6A). Hopping between intermediate states ceases when the G-domain adopts its folded structure ("G_N-II_U", Figure V-6A). Refolding of domain II is not observed under these conditions (3.5 pN, time cutoff: 120 seconds).

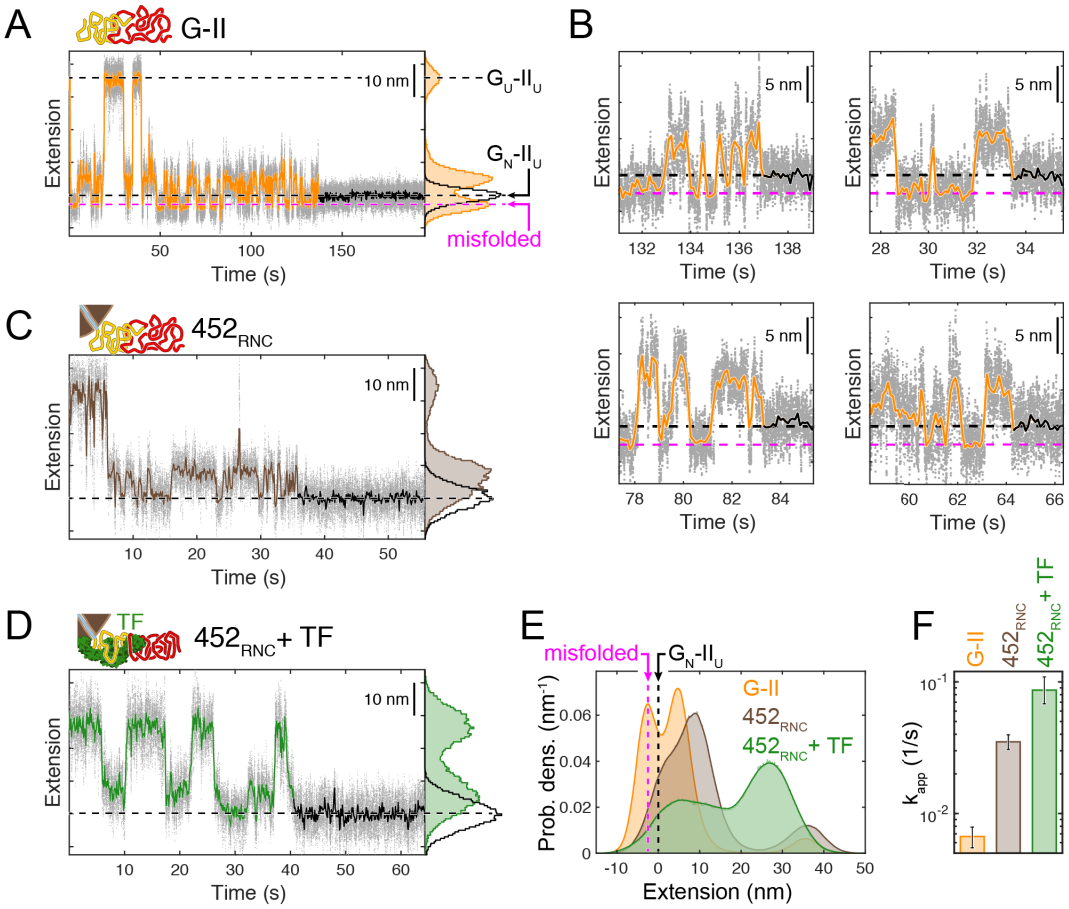


Figure V-6 The ribosome and trigger factor reduce inter-domain misfolding. (A) Example of G-domain refolding against a constant force of 3.5 pN for *G-II*. Raw data (1 kHz) is shown as grey dots, time-averaged data (10 Hz) as a line (orange before G-domain folding, black thereafter). Black dashed lines indicate states before and after G-domain folding (G_u-II_u, G_N-II_u). *G-II* also populated before (orange) and after (black) G-domain folding. (B) Close-up view of the refolding transitions of four G-II example traces. In all cases, the misfolded state partially unfolds before productive folding occurs (black and magenta dashed lines as in A). (C) Constant force refolding of 452_{RNC}. The population of the compact misfolded species is reduced, as apparent from the extension vs time trace and the extension histogram. (D) Refolding in the presence of TF. The chaperone shifts the population away from the misfolded state. (E) Aggregated extension histograms for the equilibrium portion of constant fore refolding traces, aligned to the G_N-II_u state. The misfolded state is highly populated in *G-II* (orange), but less so in 452RNC without (brown) and with (green) TF. (F) Apparent G-domain refolding rate for G-II, 452_{RNC} with and without TF. Reduced misfolding in the presence of the ribosome and the chaperone results in faster folding.

Notably, the extension of one *G-II* folding intermediate (magenta arrow, Figure V-6A) is shorter than that of G_N-II_U. This intermediate must therefore contain parts of both domains (fig. S6, and Supplementary Information). Consistent with this interpretation, the state is not observed with the G_{alone} construct (Figure V-7). Because the short intermediate does not transition directly

to the G_N-II_U state (Figure V-6B), it most likely represents a misfolded, off-pathway species. The force clamp measurements thus directly reveal inter-domain misfolding in *G-II* that interferes with G-domain folding.

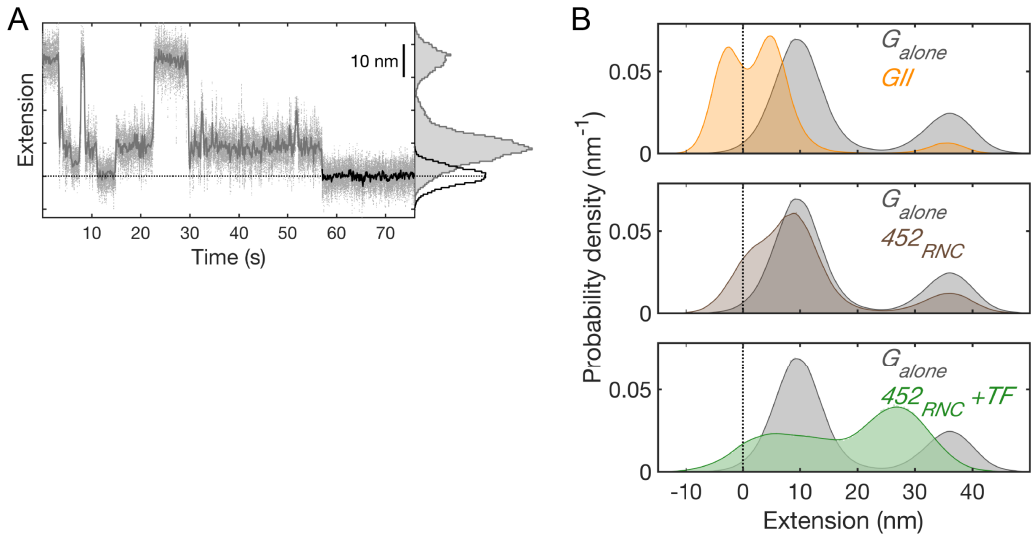


Figure V-7 Constant force data comparison among G_{alone}, *G-II* and 452_{RNC}. (A) Representative example force clamp trace of G_{alone} at 3.5 pN. Raw data (1000 Hz) is plotted as light grey dots, lines show time-averaged data (10 Hz). The equilibrium portion after refolding is in black. The histogram on the right shows the extension data for the segments before (grey) and after (black) G-domain folding. (B) Probability densities from aggregated traces for *G-II* (orange), 452_{RNC} with (brown) and without TF (green), with data for G_{alone} shown in grey for comparison. The dotted black lines in (A) and (B) indicate the extension of the state in which the G-domain is folded, which was used as a reference for aligning traces and set to 0 nm.

Remarkably, the short misfolded state that is prominently populated in *G-II* is not apparent in recordings with 452_{RNC} nascent chains (Figure V-6C). Its population is indeed much reduced in the nascent polypeptide compared to *G-II* (Figure V-6E), indicating that the ribosome reduces the propensity of domain II to form misfolded states with the G-domain. As a result, the rate of G-domain folding is higher in 452_{RNC} compared to *G-II* (Figure V-6F). The folding rate is not restored to its intrinsic value, suggesting that the nascent polypeptide still populates off-pathway species, but to a lesser degree. Hence, the ribosome appears to alleviate some of the frustration that arises when both domains navigate the folding landscape simultaneously. As

such, ribosome-nascent chain interactions may constitute a first line of defense against inter-domain misfolding.

I next assessed how the ribosome-binding chaperone TF might influence folding and misfolding of nascent EF-G. TF binds to a specific site on the ribosome near the polypeptide exit tunnel with moderate affinity (Kramer et al., 2002), and more strongly to RNCs (Kaiser et al., 2011). Significantly, EF-G has been demonstrated to be a *bona fide* TF substrate (Deuerling et al., 2003). Measurements with 452_{RNC} in the presence of TF revealed a pronounced shift in the extensions of the intermediates populated during *G-II* folding (Figure V-6D, V-6E). The chaperone causes an overall compaction of the unfolded protein and shifts the intermediate extensions away from the misfolded state more strongly than the ribosome (Figure V-6E). This remodeling of the folding landscape results in a G-domain folding rate of 0.087 s^{-1} (ci: 0.068 s^{-1}, 0.109 s^{-1}) (Figure V-6F), higher than that observed in the absence of the chaperone and closer to the intrinsic rate of G-domain folding in isolation. TF therefore appears to sequester the unfolded domain II, preventing it from inducing misfolding and thereby neutralizing its effect on G-domain folding rate. Our measurements thus show how the chaperone promotes folding of an authentic client protein.

V-2.4: Contacts with the G-domain stabilize domain II

In 452_{RNC}, the nascent chain is sufficiently long for domain II to emerge from the ribosome and fold. Indeed, the initial pull on 452_{RNC} reveals sequential unfolding transitions of the G-domain and domain II (Figure V-8). A very similar unfolding pattern is observed for *G-II* (Figure V-9A). In both cases, each domain unfolds with a characteristic pattern of either one single step or two steps. This apparent heterogeneity is explained by short lifetimes of the

unfolding intermediate (Figure V-9B), which consequently is resolved in only a fraction of events at our temporal resolution of 1 ms. Very similar transitions are observed during the initial unfolding of full-length EF-G (Liu et al., 2017), indicating that the two domains adopt the same structure in *G-II* and 452_{RNC} as in the full-length protein (Figure V-10).

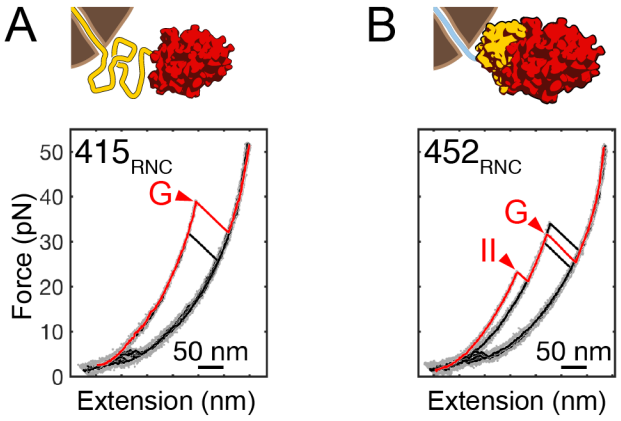


Figure V-8 Domain II folds until it is out of ribosomal exit tunnel. Example FECs for 415_{RNC} (A) and 452_{RNC} (B). The initial unfolding trace (red) shows the signature of G-domain unfolding in both constructs ("G", arrowheads). The signature of domain II ("II", arrowhead) is apparent in 452_{RNC}, but no transitions that can be attributed to domain II is apparent in 415_{RNC}.

After complete unfolding, the G-domain refolds sporadically after the polypeptide is relaxed to 2 pN. In contrast, refolding of domain II does not occur under these conditions (Figure V-9A).

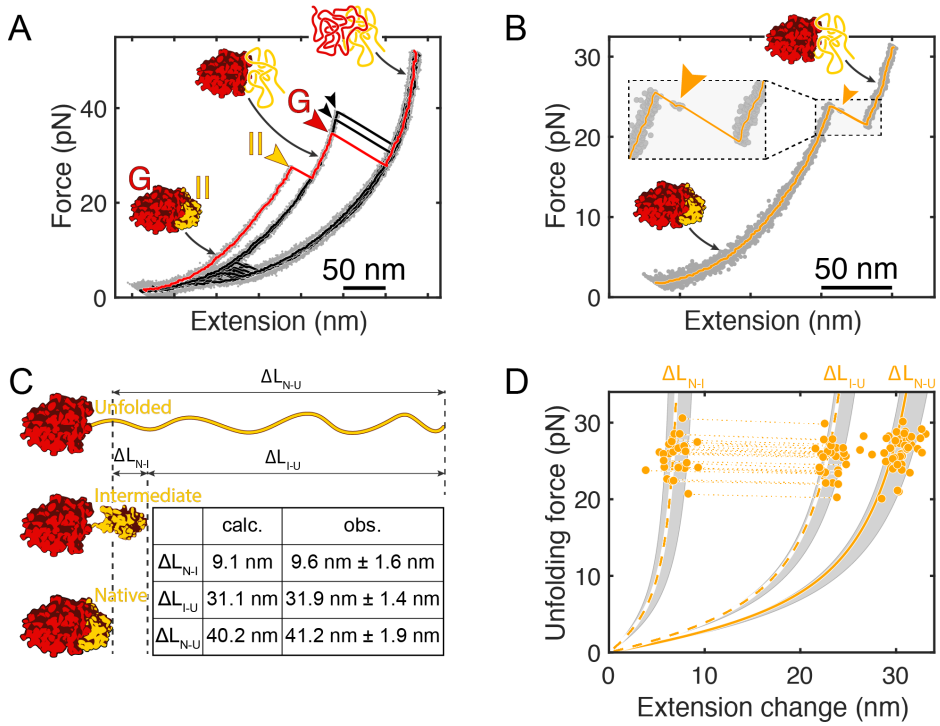


Figure V-9 Domain II unfolds via an intermediate state in *G-II*. (A) Example FECs for *G-II*. Clear signatures of both domains (colored arrowheads). Are apparent in the initial trace (red). The G-domain refolds in some of the attempts (black arrowheads). Refolding of domain II is not observed. (B) Transient population of an unfolding intermediate (arrowhead) during II unfolding. Inset: magnified view of unfolding transitions. The short lifetime of the intermediate indicates that it is unstable. (C) Schematic illustrating the likely pathway for two-step unfolding of domain II in *G-II*. The interface between the G-domain and domain II breaks first, followed by unfolding of the remaining domain II structure. The table lists the contour length changes expected based on the EF-G structure ("calc.") and the observed values ("obs") for the transitions. (D) Scatter plot of domain II unfolding transitions in *G-II*. Each dot represents an unfolding transition. Transitions from two-step unfolding via the intermediate are connected by dotted lines. Dashed and solid lines represent WLC models. Calculated for two-step ($\Delta L_{N\text{-}I}$, $\Delta L_{I\text{-}U}$) and one-step ($\Delta L_{N\text{-}U}$) unfolding. The shaded areas indicate the representative standard deviations of the observed contour length changes.

In the folded state, the G-domain and domain II form an extensive interface with a buried surface area of almost 2000 Å^2 (Pulk and Cate, 2013). These contacts conceivably contribute to the structural stability of domain II. I speculated that this interface may break when tension is applied during force ramp experiments. In such a scenario, the two domains would reorient under the applied tension, and a largely unstructured region connecting them (aa 286 – 317) would be stretched, resulting in a calculated contour length change of $\Delta L_{N\text{-}I}$(calc.) = 9.1 nm (Figure V-9C). Subsequently, unfolding of the remaining domain II structure would yield a further increase of

$\Delta L_{I\text{-}U}$(calc.) = 31.1 nm. The experimentally observed length changes ($\Delta L_{N\text{-}I}$(obs.) = 9.6 nm ± 1.6 nm, $\Delta L_{I\text{-}U}$(obs.) = 31.9 nm ± 1.4 nm; mean ± standard deviation; Figure V-9D) are very similar to these values, suggesting that domain II unfolds through this pathway in our experiments. Rapid unfolding after interface rupture indicates that domain II is unstable by itself, whereas the G-domain remains folded until the force is further increased (Figure V-9). Contacts with the folded G-domain therefore appear to be important for the stability of domain II.

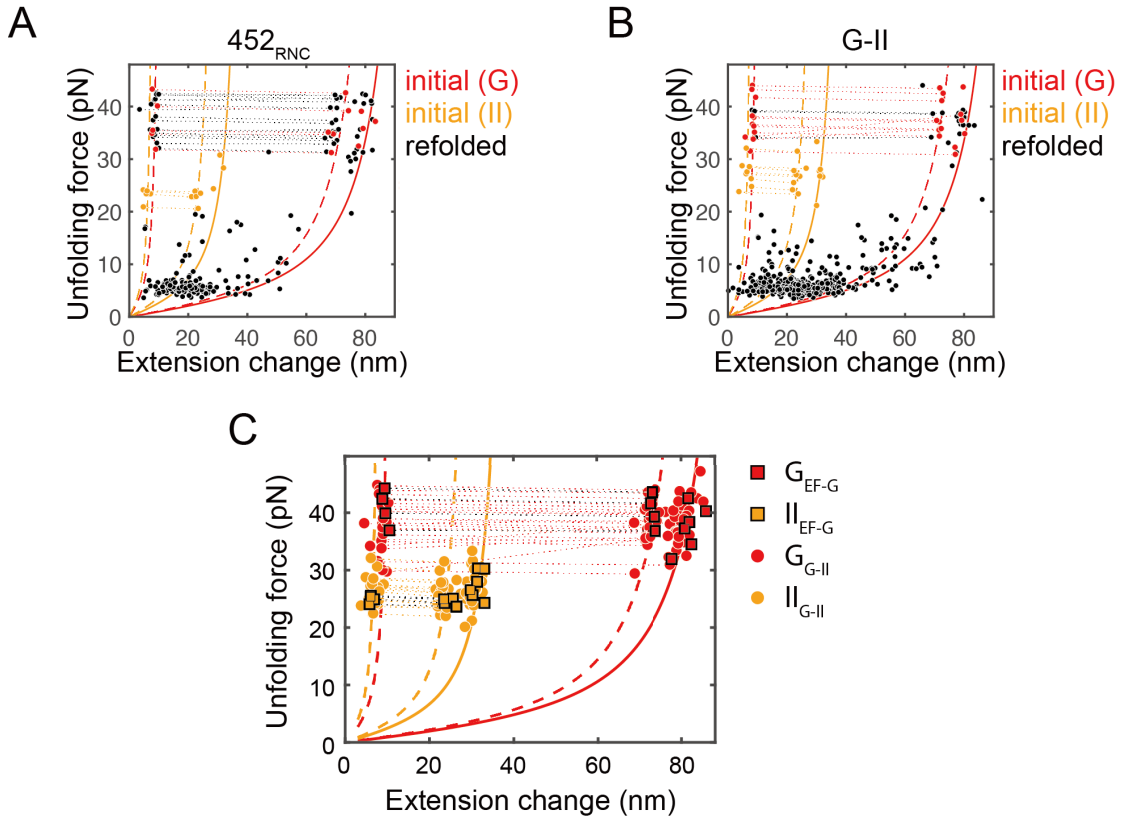


Figure V-10 *G-II* refolds to its native state both on and off ribosome. (A) Scatter plot showing unfolding transitions from *452*$_{RNC}$. Initial transitions are shown in red (G-domain) and yellow (domain II), while transitions after refolding are shown in black. Pairs of transitions from two-step unfolding are connected by dotted lines. Most of the transitions after refolding do not resemble the initial transitions and likely represent partially structured or molten globule-like states. The G-domain refolds only sporadically, and unfolding of domain II is not observed. Lines represent WLC models for G-domain (red) and domain II (yellow) unfolding (solid lines: one-step unfolding, dashed lines: two-step unfolding). (B) Scatter plot showing unfolding transitions from *G-II*. Colors and lines as in panel A. As with *452*$_{RNC}$, refolding of domain II is not observed, and refolding of the G-domain occurs only rarely, consistent with its low folding rate in this construct. (C) Scatter plot comparing initial unfolding transitions from *G-II* and full-length EF-G. The unfolding transitions of the G-domain (red) and domain II (yellow) are very similar for

both constructs. Additional transitions resulting from unfolding of domains III, IV and V that are apparent during unfolding of full-length EF-G are not shown here. The red curves represent WLC models for unfolding of the native G-domain, yellow curves represent WLC models for domain II unfolding (solid lines: one-step unfolding, dashed lines: two-step unfolding). Pairs of transitions from two-step unfolding are connected by dotted lines.

V-2.5: Domain II folding requires the folded G-domain

We hypothesized that the dependence of domain II stability on the interface with the folded G-domain might impose a sequential order on domain folding during EF-G synthesis: domain II can stably fold only after the G-domain has acquired its native structure. I used the *G-II* construct to test whether folding of domain II ($II_{G\text{-}II}$) indeed requires the folded G-domain ($G_{G\text{-}II}$), exploiting the differential mechanical stability of the two domains (Figure V-11). Domain II always unfolds before the G-domain in force ramp experiments, providing an opportunity to assess the importance of native interface contacts for folding.

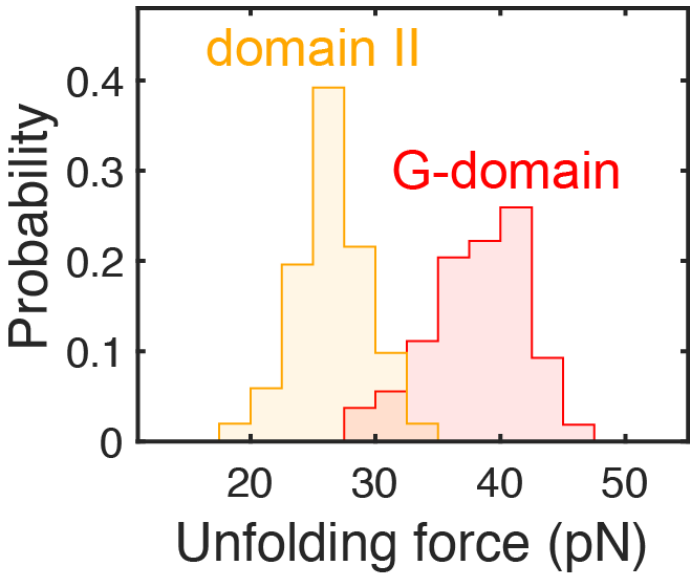


Figure V-11 Unfolding force distributions of the G-domain and domain II. Comparison of unfolding force distributions from initial domain II (yellow) and G-domain (red) unfolding of *G-II*. The G-domain unfolding forces are largely separated from those of domain II, with domain II being mechanically weaker and always unfolding first in force ramp experiments.

I selectively unfolded $II_{G\text{-}II}$ in force ramp experiments, keeping $G_{G\text{-}II}$ folded, before lowering the force to initiate refolding (Figure V-12A). Domain II refolds under these

conditions, typically after several force ramp cycles (Figure V-12A). Unfolding transitions after refolding are indistinguishable from the initial ones (Figure V-12B), indicating that $II_{G\text{-}II}$ regained its native structure under these conditions. Contacts with the natively folded G-domain therefore appear to be required for stable structuring of domain II. Consistent with this notion, a polypeptide containing only domain II did not accumulate upon expression of a recombinant construct in *E. coli* (data now shown), presumably because the translation product is not stably structured and quickly degraded by cellular proteases. Our results imply a requirement for ordered domain-wise folding: $II_{G\text{-}II}$ folds only after $G_{G\text{-}II}$ has folded (Figure V-12C).

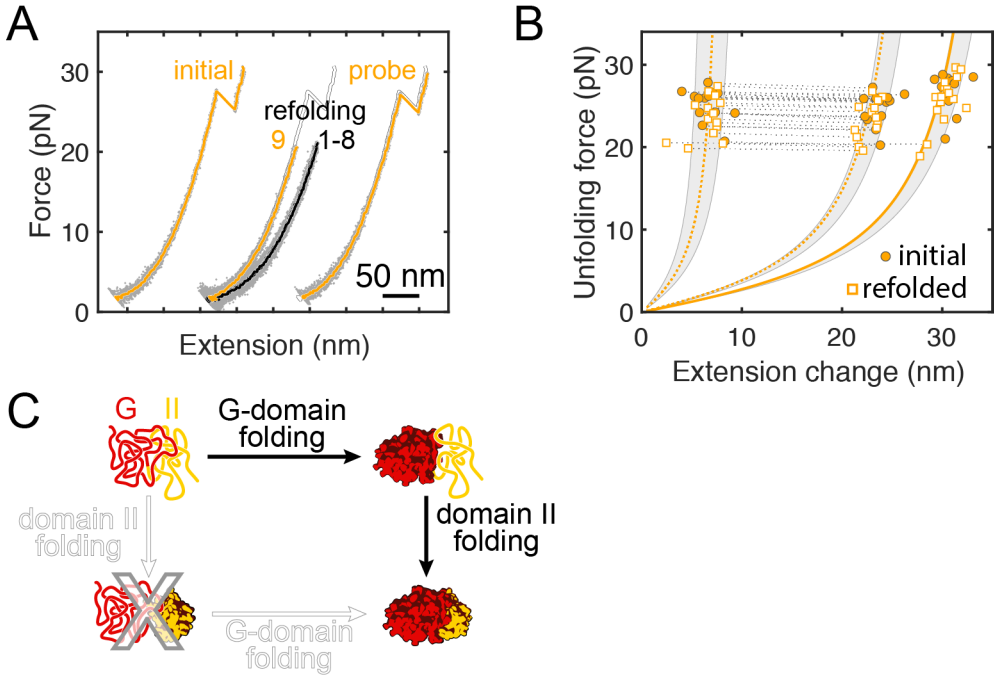


Figure V-12 Domain II folding requires the folded G-domain. (A) Example FECs showing slow folding of domain II. Domain II is selectively unfolded while leaving the G-domain structured (yellow trace, "initial"). Refolding is apparent in trace "9" (yellow) after several futile attempts (black traces, "1-8"). Subsequent forced unfolding confirms domain II refolding (yellow trace, "probe"). Traces are plotted with an offset along the x axis for clarity, the "initial" trace is re-plotted in white as a reference. (B) Scatter plot for unfolding of renatured domain II, confirming folding to the native structure in the presence of the folded G-domain. Dashed and solid yellow lines represent WLC models for two-step and one-step unfolding, shaded regions indicate the standard deviation of the experimental data. (C) Domain II does not stably fold by itself because it requires stabilizing contacts at the interface with the G-domain. Consequently, the G-domain has to fold first, imposing a hierarchical folding order.

V-2.6: Unfolded domain II destabilizes the native G-domain

Strikingly, refolding of domain II was not the only outcome we observed after selectively unfolding $II_{G\text{-}II}$. In some refolding attempts, $G_{G\text{-}II}$ unfolded at uncharacteristically low forces, often in the range between 5 and 10 pN (Figure V-13A and B). In contrast, unfolding of native $G_{G\text{-}II}$ invariably occurred between ~30 and 50 pN (Figure V-13B), similar to G_{alone} and all nascent chains. The low force transitions are well outside this range (Figure V-13B), indicating that $G_{G\text{-}II}$ had assumed a non-native conformation before we probe its structure mechanically. The transitions resemble partially folded and molten globule-like states populated by 328_{RNC} (Figure V-13D) and *G-II* (Figure V-13E) before G-domain refolding, suggesting that $G_{G\text{-}II}$ progressively loses its native structure and subsequently enters into misfolding interactions with domain II. Low-force unfolding is not observed with G_{alone} (0 events in 399 trials, p-value: $1.1\text{x}10^{-27}$, see Materials and Methods), confirming that interactions with the unstructured domain II pull the G-domain out of its native state. It thus appears that unfolded $II_{G\text{-}II}$ "denatures" $G_{G\text{-}II}$, leading to a complete loss of nativeness (Figure V-13D).

During EF-G synthesis, the G-domain is likely to fold while domain II is being translated, as described above. Our experiments indicate that the polypeptide can then proceed along one of two routes, leading to either folding of domain II (Figure V-12) after its complete synthesis, or to denaturation of the G-domain (Figure V-13) and subsequent misfolding. To determine which outcome is more likely, we compared the cumulative probabilities of the two processes (Figure V-13C). G-domain denaturation and domain II folding occur on similar timescales in the *G-II* construct. This observation suggests that folding and denaturation compete with each other in the isolated *G-II* polypeptide (Figure V-13F).

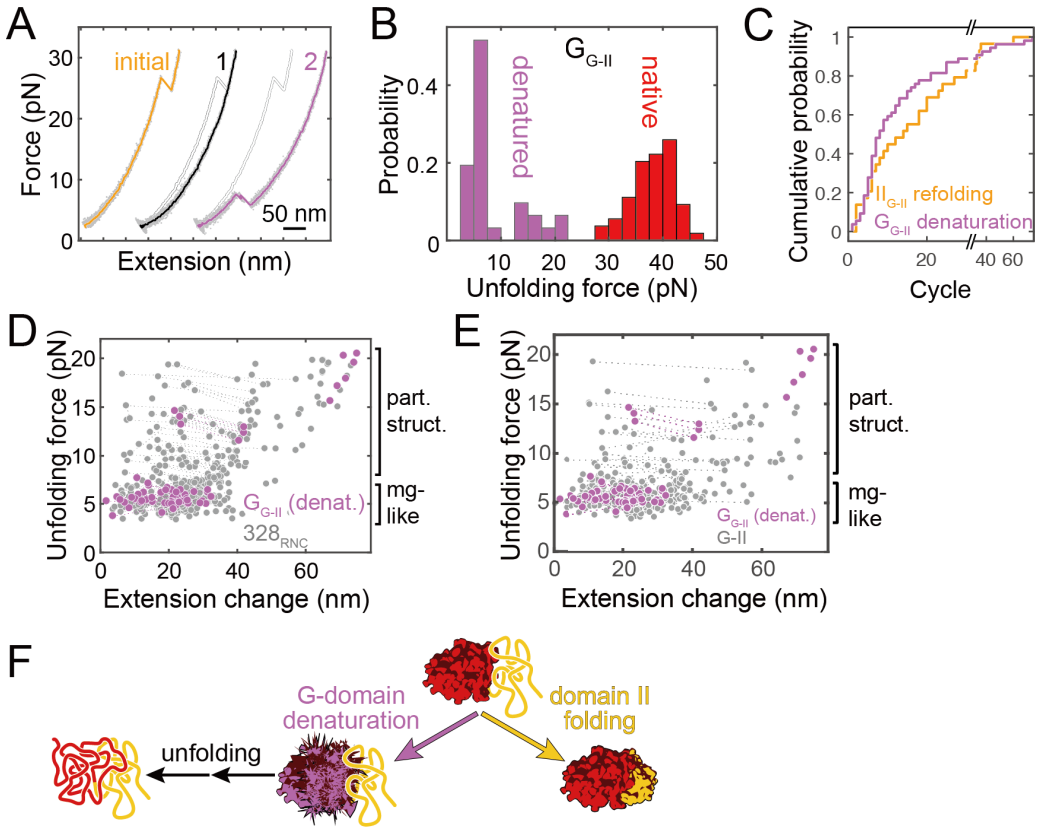


Figure V-13 Unfolded domain II destabilizes the native G-domain. (A) Example of G-domain denaturation. After selectively unfolding of domain II (yellow, "initial"), unfolding of the G-domain at low force during the second cycle (magenta, "2") indicates denaturation. The initial trace is re-plotted in white as a reference. (B) Distribution of unfolding forces. The native G-domain (red) unfolds in a force range that is well separated from that after denaturation (magenta). (C) Scatter plot showing individual unfolding events after G-domain denaturation (magenta). The distribution resembles that of molten globule-like and partially structured states observed after mechanical unfolding of *328*$_{RNC}$ (grey dots), indicating that denaturation ultimately results in complete unfolding. (D) Cumulative probability of domain II refolding (yellow) and G-domain denaturation (magenta) during repeated force ramp cycles. Both events occur with similar probability and thus on similar timescales. (E) In addition to folding productively, domain II can cause G-domain denaturation, which then results in complete unfolding of the polypeptide. The two processes compete with each other.

V-2.7: TF protects against denaturation

To determine whether the ribosome protects the folded G-domain against denaturation by unfolded domain II, I repeated the experiment with the *452*$_{RNC}$ construct. In contrast to the frustration resulting from interactions among unfolded domains, denaturation of natively folded $G_{G\text{-}II}$ is not alleviated by the ribosome (Figure V-14A). Consequently, even after the G-domain in

nascent EF-G has folded, it is still at risk of losing its native structure as long as domain II is unfolded. Subsequently, the two non-native domains likely enter into misfolding, as observed after complete unfolding of *G-II*. Denaturation of the G-domain therefore competes with folding of domain II (Figure V-14B) and thus has the potential to derail productive cotranslational folding.

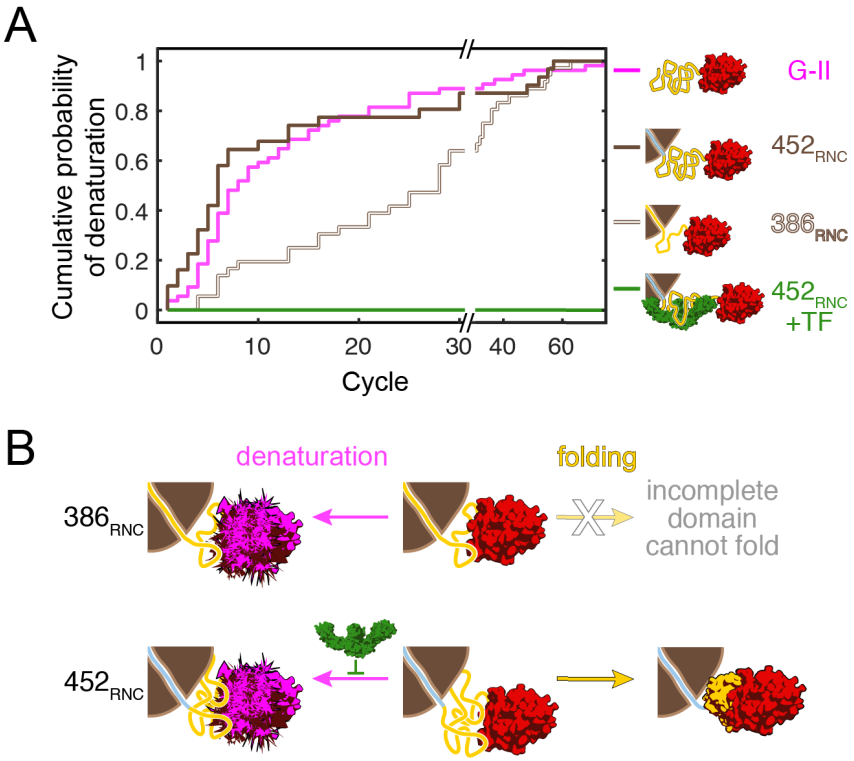


Figure V-14 TF protects against denaturation. (A) Cumulative probability distributions of G-domain denaturation. The ribosome does not protect against denaturation (*G-II* vs *452*$_{RNC}$), which already occurs after synthesis of domain II (*386*$_{RNC}$). TF blocks denaturation (*452*$_{RNC}$ + TF). (B) Before domain II has been fully synthesized (*386*$_{RNC}$), it cannot fold, but already destabilizes the neighboring folded G-domain. After domain II has emerged from the ribosome *(452*$_{RNC}$), competing pathways result in either G-domain denaturation or folding of domain II. TF blocks denaturation, favoring folding.

Because folding of protein domains is cooperative, domain II can adopt its native structure only after it has fully emerged from the ribosome. Consistent with this notion, we did not observe any stable structures attributable to domain II in experiments with *415*$_{RNC}$ (Figure V-8A) or shorter nascent chains. Denaturation of the G-domain, on the other hand, is already

observed when only 386 codons have been translated (386_{RNC}, Figure V-14A). In 386_{RNC}, only about half of domain II has exited the ribosome exit tunnel, but destabilization of the folded G-domain structure is already observed (Figure V-14A), and constructs containing truncated domain II aggregated upon expression in *E. coli* (data not shown). These results suggest that denaturation can begin before productive folding is possible. In the context of elongating nascent EF-G, the balance between productive folding and denaturation is therefore skewed toward the non-productive pathway (Figure V-14B).

Denaturation of the G-domain by unfolded, elongating domain II represents an unanticipated complication of domain-wise cotranslational folding. How is the nascent polypeptide kept on a productive folding track in the cell? As shown above, the ribosome does not protect against denaturation. In contrast, TF completely prevented denaturation of the G-domain (Figure V-14A). Given a base probability of denaturation of $p = 0.144$ per cycle (Materials and methods), the probability that we missed similar denaturation events in the presence of TF is very low (0 events in 167 trials, p-value from a binomial distribution test: 5.3×10^{-12}). Our experiments thus reveal a dual role of TF during early EF-G folding: The chaperone not only helps to resolve the frustration resulting from inter-domain misfolding (Figure V-6), but also protects the folded G-domain against denaturation by the unfolded domain II (Figure V-14).

V-3: Discussion

The early events during the folding of a nascent multidomain protein, EF-G were defined (summarized in Figure V-15). Interactions of the nascent polypeptide with the ribosome are beneficial for folding soon, but not immediately after the domain has been synthesized (Figure

V-5). Experiments with full-length EF-G (Liu et al., 2017) and other multidomain proteins (Jahn et al., 2016; Mashaghi et al., 2014; Scholl et al., 2017) have previously suggested misfolding among domains. Here, misfolding among EF-G domains was directly observed (Figure V-6), and further demonstrate that a natively structured N-terminal G-domain is a prerequisite for stable folding of the subsequently synthesized domain II (Figure V-9 and V-12). In the case of EF-G, domain folding is therefore hierarchical and occurs in the order in which domains are synthesized, promoting productive cotranslational folding. Failure of the G-domain to fold soon after domain II begins to emerge likely has deleterious consequences for overall EF-G folding. As more and more unfolded polypeptide accumulates during synthesis, misfolding makes completion of the required early folding steps increasingly difficult, resulting in global misfolding and irreversible aggregation. Folding of the N-terminal domain therefore constitutes an important waypoint along the folding pathway.

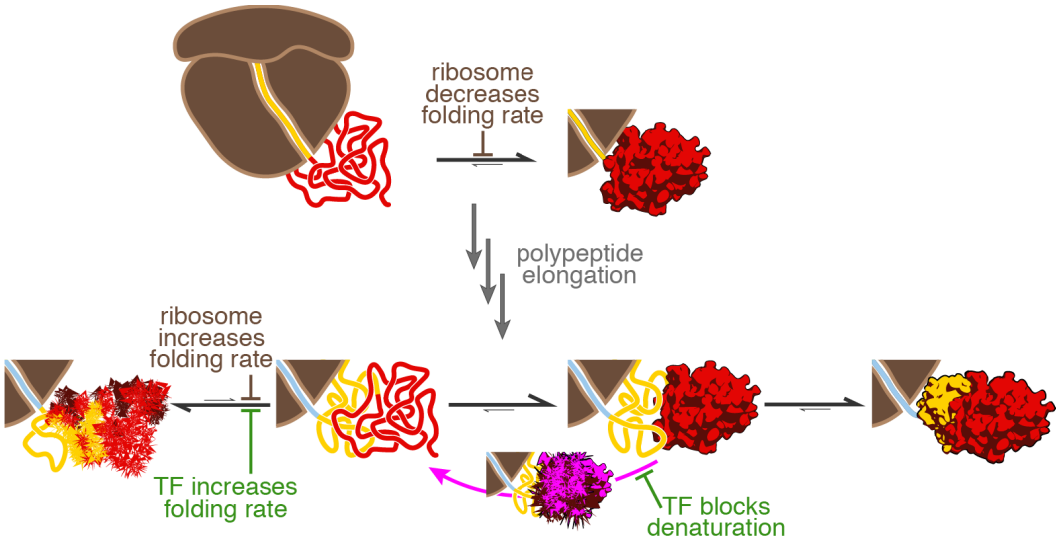


Figure V-15 Folding events during EF-G synthesis. After the G-domain has emerged from the ribosome (top), its folding rate is reduced by interactions with the ribosome. Presumably similar interactions, now with domain II, reduces inter-domain misfolding in longer nascent chains (bottom), effectively increasing the G-domain folding rate. How the ribosome affects a particular domain thus depends on nascent chain length. The reduction in inter-domain misfolding is reinforced by TF. G-domain folding has to precede folding of domain II. Domain II folding competes

with denaturation of the already folded G-domain. The latter process is prevented by TF, but not the ribosome. Hence, TF has a dual function in promoting early folding events in nascent EF-G.

Previous studies with small proteins artificially tethered to the ribosome (Hoffmann et al., 2012; Kaiser et al., 2011; Samelson et al., 2016), have suggested that the ribosome helps to prevent premature folding by destabilizing folded structures and reducing folding rates. Here, a natural multidomain protein also shows this effect (Figure V-5). However, we observe an additional consequence of ribosome-nascent chain contacts that is potentially more important for multidomain proteins (Figure V-6): sequestering the proximal part of the nascent chain (domain II in 452_{RNC}) reduces the formation of misfolded states with more distal parts (the G-domain in 452_{RNC}). The effect of the ribosome on overall folding is therefore highly dependent on nascent chain length and can manifest as either decelerated (328_{RNC} *vs.* G_{alone}) or accelerated (452_{RNC} *vs. G-II*) folding (Figure V-5). These findings provide a framework for studying the bioinformatically (Chaney et al., 2017; Pechmann and Frydman, 2013) and experimentally (Buhr et al., 2016; Mercier and Rodnina, 2018; Zhang et al., 2009) observed tuning of synthesis and folding rates.

TF has been proposed to act as a "holdase" for nascent proteins, preventing and even reversing premature folding (Hoffmann et al., 2012). Here, the results proved that TF prevents inter-domain misfolding within nascent EF-G chains, amplifying a similar contribution from the ribosome (Figure V-6). As a result, the chaperone can effectively speed up folding (Fig. 2F), in contrast to the previously observed global deceleration (Agashe et al., 2004). Interestingly, our data suggest that the folding pathway of the G-domain changes in the presence of TF, as indicated by the population of intermediate states with altered molecular extensions (Figure V-6). A previous study, using the periplasmic maltose binding protein, suggested that TF stabilizes

partially folded structures (Mashaghi et al., 2013). A simpler explanation, consonant with a recent structural characterization of TF-substrate interactions (Saio et al., 2014), could be that the chaperone compacts the largely unfolded substrate protein by simultaneously binding to several sites. The changes in state extensions in the constant force refolding measurements are consistent with this interpretation. Another consequence of forming multiple contacts that are spaced far apart in the client protein primary structure is a reduction in chain entropy, which facilitates folding (Haldar et al., 2017).

Unexpectedly, the data revealed that an unfolded part of the nascent polypeptide can denature an already folded domain (Figure V-13). While the ribosome does not protect against denaturation, the chaperone TF effectively prevents it (Figure V-14). Therefore, in addition to preventing the misfolding of unfolded or partially structured states, the chaperone also protects already folded structures (Figure V-14). While some proteins utilize disordered sequences for function and regulation (Motlagh et al., 2016), others have evolved to remain stably folded. Intertwined domain topologies (Shank et al., 2010) and interactions among topologically separate domains (Flaugh, 2005; Han et al., 2007) may serve to minimize the population of partially unfolded states that could potentially initiate more global unfolding. Nevertheless, unfolded segments are necessarily present during protein synthesis by the ribosome for extended periods of time, due to relatively slow elongation rates (~10-20 amino acids per second in bacteria (Young and Bremer, 1976)). An unanticipated major role of nascent chain-binding chaperones may thus be to protect folded domains against denaturation by unfolded sequence segments during cotranslational folding. Whether chaperones functioning post-translationally to

maintain a folded proteome act through similar mechanisms remains to be assessed in future studies.

V-4: Materials and methods

*V-4.*1: Folding *Cycle Analysis*

To determine refolding probabilities from force ramp experiments, proteins or RNCs were subjected to repeated cycles of denaturation (pulling to the maximum force, see above) and refolding (relaxing to a force of 2 pN and holding that force for 10 seconds). After initial denaturation, the number of cycles until refolding occurred were counted (Figure V-4). From a large number (>100) of cycles, a distribution of refolding events as a function of cycle number was obtained and converted into a probability distribution. This distribution should follow a geometric distribution, assuming that folding occurs with a constant probability within any one cycle. Since we have a limited amount of data (i.e., coverage over all possible cycle numbers is finite), the following function was used to fit our data:

$$P_i = \alpha * p * (1 - p)^{i-1},$$

where P_i is the probability of waiting i cycles to observe folding, p is the folding probability within one cycle, and α is a scaling factor to account for incomplete sampling of the geometric distribution. Therefore, the cumulative distribution for the first n cycles is

$$P_n = \alpha * [1 - (1 - p)^n],$$

where P_n is the cumulative probability up to cycle n.

*V-4.*2: Folding *rate estimation*

To estimate apparent folding rates ("Apparent rate" in Figure V-5) from different proteins and RNCs from the folding probability p obtained from FCA, we used the first order approximation

$$k = \frac{-\ln(1-p)}{t},$$

where k is the first order apparent rate constant, p is the folding probability within one cycle, and t is the folding time, which equals 10 seconds in our study. In constant force refolding measurements, multiple intermediate states were observed during G-domain folding at 3.5 pN. The refolding therefore is not 2-state, and the first-order approximation represents an oversimplification. However, refolding times from force clamp measurements are well described by a single-exponential process (data not shown), indicating that the simplified assumption of a single rate-limiting step yields a reasonable estimate of the overall folding rate. In addition, the distribution of folding events along repeated force ramp cycles is well described by the geometric distribution.

*V-4.*3: G-*domain denaturation and II-folding probability estimation*

Refolding of domain II and denaturation of the G-domain by interactions with domain II are both slow processes. As a consequence, the available dataset is limited. In addition, the two types of events are mutually exclusive. We therefore modified our analysis to estimate the probabilities. Assuming that the number of either event D(n) (D_1(n) for G-domain denaturation, and D_2(n) for domain II folding) occurring after n cycles is described by a Poisson distribution, the maximum likelihood probability $P_{i_{max}}$ of either event happening within one cycle is

$$P_{i_{max}} = \frac{\mathrm{sum}(D_i(n))}{n * \mathrm{sum}\left(\sum_{i=1}^{2} D_i\,(n)\right) + N_0} = \frac{\#\mathrm{events(i)}}{\#\mathrm{total\ cycles}},$$

where #events(i) is the number of observed events of either G-domain denaturation (i=1) or domain II folding (i=2), and N_0 is total number of cycles in which we do not observe either event before tether breaking. Using this approach, we estimated that the probability of G-domain denaturation in any cycle is $P_{1_{max}} = 0.144$, and that of domain II folding is $P_{2_{max}} = 0.10$.

*V-4.*4: Interpretation *of force clamp refolding experiments*

In force clamp measurements, the unfolded polypeptide refolds against a constant force. Changes in extension, monitored over time, reflect structural transitions in the polypeptide. At the beginning of the trace shown in Figure V-6A, the polypeptide is in the unfolded state (G_U-II_U), at the end, the G-domain has folded (G_N-II_U). In addition, the molecule populates several partially folded states with extensions between these two states, as well as a state that is shorter than G_N-II_U by 2.5 nm. Because this short state does not directly transition to the G_N-II_U state, we interpret it to be a misfolded state:

$$G_U\text{-}II_U \rightleftharpoons [I] \rightharpoonup G_N\text{-}II_U \quad (I = \textit{misfolded})$$

In force clamp measurements, the observed difference in extension is the difference in end-to-end distance between the initial and the final states:

$$\Delta x_{observed} = x(\textit{initial}) - x(\textit{final})$$

In the case of the initial and final state during *G-II* refolding measurements (Figure V-16A), the observed change is

$$\Delta x_{observed} = x(G_U\text{-}II_U) - x(G_N\text{-}II_U)$$

The extension of G_U-II_U depends only on force, which is constant in this experiment. The shorter extension of the misfolded state must therefore be due to a change in the second term. The extension of G_N-II_U is the sum of the extensions of the two domains:

$$x(G_N\text{-}II_U) = x(G_N) + x(II_U)$$

The larger observed extension change for the misfolded state could in principle be due to a smaller $x(G_N)$, i.e. the G-domain adopts a state with a shorter end-to-end distance. However, the end-to-end distance of the natively folded G-domain is $x(G_N) = 0.86$ nm. A change in $x(G_N)$ therefore cannot account for an observed difference of 2.5 nm. Thus, the state with the misfolded state must involve both domains (Figure V-16B).

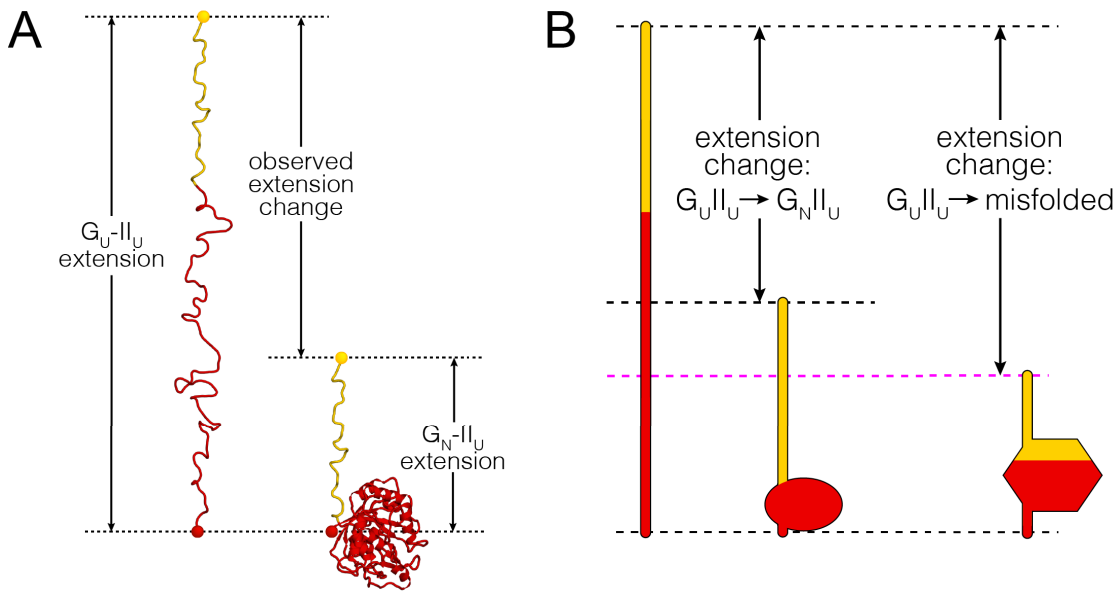


Figure V-16 Interpretation of extension changes from force clamp experiments. (A) The extension change observed during refolding at constant force is the difference in extension between the initial (fully unfolded; G_U-II_U) state and the final state of the molecule, in which the G-domain is folded (G_N-II_U). The folding intermediates observed in these experiments are not represented here. (B) A shorter extension, indicated by the magenta line, is most likely due to the formation of a structure (hexagon) that contains elements from both the G-domain and domain II.

*V-4.*5: Binomial *test*

To assess whether G_{alone} spontaneously undergoes denaturation in the absence of domain II, we conducted control experiments in which we subjected G_{alone} to the same force ramp conditions as G-II to detect possible denaturation. From total of 7 molecules with 399 trials, we observed no denaturation events with G_{alone}. A binomial test with a denaturation probability per cycle of 0.144, which is estimated from G-domain denaturation experiments performed with *G-II*, yielded a p-value of 1.1×10^{-27}. It is thus extremely unlikely that G_{alone} exhibits the denaturation observed with *G-II*. Domain II therefore is responsible for pulling the G-domain out of its native state. A similar analysis was performed to confirm that we did not miss G-domain denaturation in the presence of TF, which is also very unlikely (p-value: 5.3×10^{-12}).

*V-4.*6: Kolmogorov-*Smirnov test (K-S test)*

A one dimensional two-sample K-S test (Massey, 1951) was performed to evaluate whether initial unfolding transitions of domain II are different from those obtained after refolding. We compared the distributions of unfolding forces and contour length changes separately. In both cases, the results suggest that it is very likely that we are sampling the same underlying distribution, which indicates that the structure of domain II is the same in both cases (Figure V-17). In addition, we conducted a two-dimensional two-sample K-S test (Peacock, 1983) to compare unfolding events (each characterized by unfolding force and extension change), which confirmed that the distributions for native and refolded domain II are not significantly different (Figure V-17C).

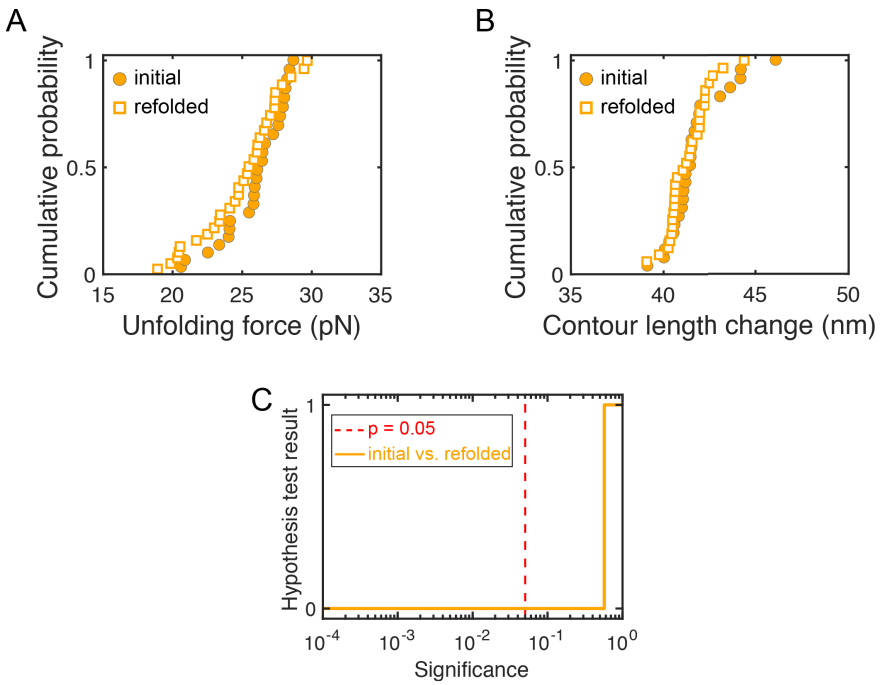


Figure V-17 Domain II refolds into its native structure. Cumulative distribution of unfolding forces (A) and contour length changes (B) of domain II from initial unfolding (circles) and unfolding after renaturation (squares). The G-domain was kept in the folded state by limiting the maximum force to 20 pN. The similarity of the distributions indicates that domain II regains its native structure. This result is confirmed by one-dimensional two-sample Kolmogorov-Smimov tests (Massey, 1951), which indicate that the respective data sets are from the same underlying dirstibution (p = 0.05). (C) Rersults from two-dimensional two-sample Kolmogorov-Smimov test (Peacock, 1983), comparing unfolding transitions of initial and refolded domain II unfolding. A test result of 0 indicates that the null hypothesis ("both data sets are drawn from the same distribution") cannot be rejected. The test was evaluated over a range of significance levels, and the result indicates that the two distributions are not significantly different (p = 0.57).

The two-dimensional two-sample K-S test was also performed to evaluate whether the transitions from denatured G-domain are different from those of *G-II* or 328_{RNC} (Figure V-18). The results suggest that the distributions for *G-II* and 328_{RNC} are distinct, but neither one is different from that of the denatured G-domain. This observation suggests that denaturation of the G-domain leads to the population of molten globule-like and partially folded states (similar to those populated during 328_{RNC} refolding) and G-domain/domain II misfolded states (similar to those populated during *G-II* refolding after complete mechanical unfolding). It therefore appears

that denaturation of the G-domain by domain II leads to complete unfolding via the intermediates that are observed during refolding of 328_{RNC}.

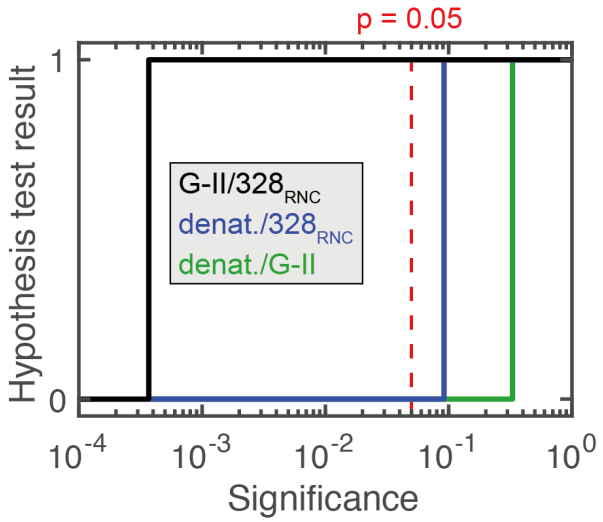


Figure V-18 The denatured G-domain resembles folding intermediates of G-domain. Results of pairwise two-dimensional Kolmogorov-Simimov tests comparing distributions of unfolding events evaluated at different significance levels. A test result of 0 indicates that the null hypothesis ("data sets are from the same underlying distribution") cannot be rejected. Only unfolding events from non-native transitions are compared. The test indicates that events from G-II and 328NC are significantly different, whereas neither distribution differs significantly form unfolding of the denatured G-domain. This result indicates that unfolding transitions of the denatured G-domain have significantly similarity to those of 328_{RNC} and *G-II*, and suggests that denaturation results in the population of partially structured states and, ultimately, complete unfolding.

Chapter VI: Energetics of domain-domain interactions in a complex multidomain protein

By

Kaixian Liu

This work is to be submitted as:

Liu, K.*, Chen, X.*, Kaiser, C.M. (2019). Energetics of domain-domain interactions in a complex protein. (In preparation)

*: These authors contributed equally to this work.

Contributions: Kaixian Liu purified EF-G and *G-II-III*, and Xiuqi Chen purified *III-IV-V*. Kaixian Liu and Xiuqi Chen together collected single-molecule data. All authors analyzed the data, created figures, and wrote the manuscript.

Chapter VI. Energetics of domain-domain interactions in a complex multidomain protein

VI-1: Introduction

Multidomain proteins constitute a large fraction of all proteomes, but studies investigating how these complex proteins fold into their functional structures are scant (Han et al., 2007). Many multidomain proteins are prone to aggregation in ensemble folding measurements, complicating mechanistic folding studies. Most multidomain protein folding studies to date have focused either on tandem repeat proteins (Batey and Clarke, 2008; Batey et al., 2005, 2006; Scott et al., 2002) or relatively simple multidomain proteins (Parker et al., 1996; Rudolph et al., 1990; Sánchez et al., 2004). How domain-domain interactions contribute to the folding and stability of natural multidomain protein has remained elusive. Recently, single-molecule fluorescence (Borgia et al., 2015, 2011) and force (Bauer et al., 2015; Jahn et al., 2018; Liu et al., 2017, 2019; Scholl et al., 2014) spectroscopy techniques have begun to shed light on how domain interactions within multidomain proteins affect protein folding mechanisms and stabilities. Single-molecule approaches are particularly well suited to resolve the heterogeneity intrinsic to multidomain protein folding pathways. Previous studies have shown that simultaneous folding of several domains results in a high propensity to form misfolded off-pathway states, which hampers productive folding (Borgia et al., 2011; Jahn et al., 2016; Liu et al., 2019). The notion that cotranslational domain-wise folding alleviates folding frustration is widely accepted, but few detailed studies are available to date (Evans et al., 2008; Fedorov and Baldwin, 1997; Frydman et al., 1999; Netzer and Hartl, 1997).

Once a multidomain protein reaches its native state, domains typically form numerous interactions that increase the overall stability of the native structure (Marsh and Teichmann, 2014). It has been suggested that the stabilities of inter-domain interactions are highly correlated with buried surface area (Chen et al., 2013; Horton and Lewis, 1991). However, it remains unclear whether this correlation is generally applicable to predict stability from the amount of buried surface area.

Here, we have dissected the energetic contributions of neighboring domains to the stability of domain III of elongation factor G (EF-G, UniProtKB P0A6MB, Figure VI-1), a five-domain protein with an essential function in protein synthesis. Interestingly, domains III, IV and V of EF-G structurally mimic the anticodon stem loop of tRNAs in ternary complexes, highlighting the importance of domain architecture to the function of EF-G (Yamamoto et al., 2014). The poor resolution of domain III in the electron density maps of the first EF-G crystal structures suggests an intrinsic flexibility of domain III (Czworkowski et al., 1994), which may be functionally important for EF-G to transition between compact and elongated conformations during the translation elongation process (Lin et al., 2015; Pulk and Cate, 2013). Using single-molecule force spectroscopy, we extracted equilibrium and non-equilibrium parameters (Hummer and Szabo, 2001) of domain III folding in the presence and absence of its neighboring domains. We found that thermodynamic stability of domain III relies on native interactions with the C terminal domains IV and V. Surprisingly, the N terminal domains G and II do not stabilize domain III, even though domain III forms similarly extensive contacts with the N-terminal and the C-terminal domains. These results demonstrate that buried surface area alone is not a reliable predictor of energetic coupling. Additionally, the data also reveals that domain III can adopt a

stable structure only after the complete protein has been synthesized, suggesting a post-translational folding mode, which is consistent with the conclusion that eEF-2, a homolog of EF-G in eukaryotes, is likely to fold post-translationally (Mönkemeyer et al., 2019). This observation also contrasts with folding of the two N-terminal domains for which sequential co-translational folding occurs in the order of synthesis (Liu et al., 2019). Therefore, we conclude the de novo folding mechanism of a complex protein is determined by the energetic dependencies among domains.

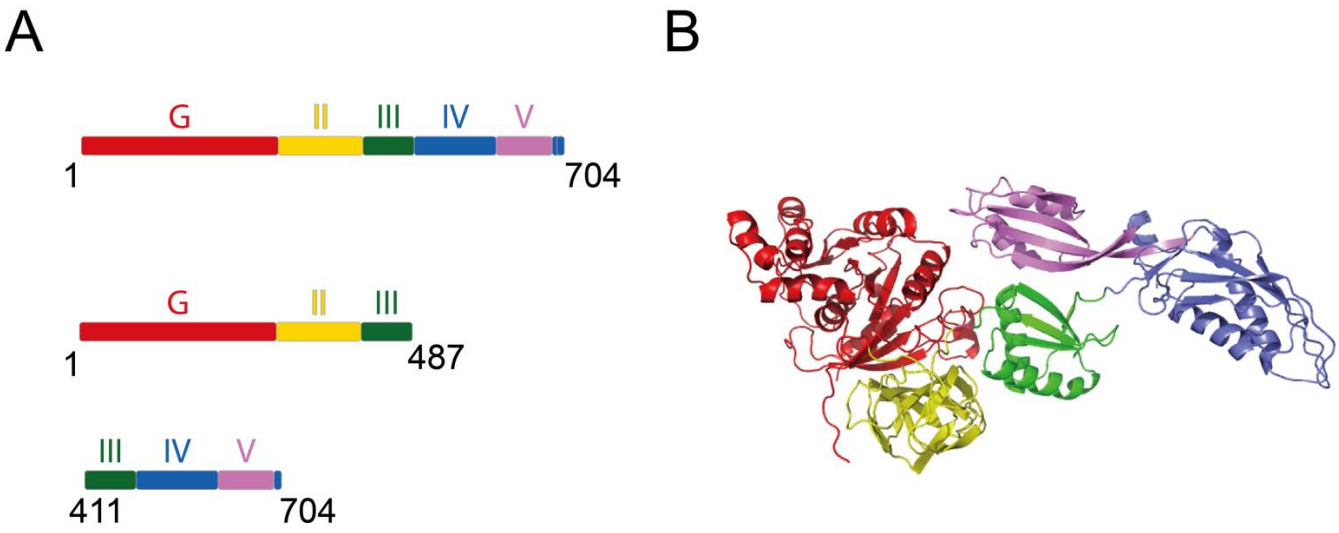


Figure VI-1 Structure diagrams of EF-G and its truncations. (A) Schematic diagram of primary structures of the three constructs used in this study, which are full-length EF-G, *G-II-III* and *III-IV-V*. Individual domains are colored as before (Figure IV-9A). (B) Cartoon representation of the crystal structure of EF-G (pdb 4V9P). Coloring scheme as in (A).

VI-2: Results

VI-2.1: Domain III is the mechanically weakest domain in full-length EF-G

We have previously characterized the unfolding of individual EF-G domains using single-molecule force spectroscopy (Liu et al., 2017), and mapped unfolding transitions of individual domains to the EF-G crystal structure (Figure VI-1). While transitions for domains G,

II and IV that were observed upon EF-G unfolding (Figure VI-2A and B) could be unambiguously assigned, the assignment of unfolding transitions for domains III and V was ambiguous, because these two domains may yield very similar length changes upon mechanical unfolding (Table VI-1). However, the topological coupling of domains IV and V (domain IV is split in primary structure into IVa and IVb, which bracket domain V) and the unfolding force distributions suggested that the initial unfolding transitions, occurring mostly below 10 pN, reflected unfolding of domain III (Figure VI-2). Analysis of unfolding transitions from full-length EF-G as well as truncated variants confirmed this assignment (Table VI-2).

domain	aa	$\Delta L_{C,calc}$ (nm)
III	411-485	24.5
IVa/b	486-614/ 687-704	45/4.5
V	615-686	25.3

Table VI-1 Comparison of contour length changes for unfolding of individual domains. The contour length changes are calculated based on and the crystal structure (Figure VI-1B).

domain	aa	Δx_{native} (nm)	$\Delta L_{C,calc}$ (nm)	$\Delta L_{EF\text{-}G}$ (nm)	$\Delta L_{III\text{-}IV\text{-}V}$ (nm)
III	411-485	2.5	24.5	26.5±2.5	25.3±2.9
IVa	492-614	0.5	43.4	45.5±2.9	44.7±1.8
IVb + V	615-704	1.6	30.8	29.3±2.8	29.5±1.8

Table VI-2 Comparison of calculated and experimental contour length changes. "aa" is the residue range for each domain. "Δx_{native}" is the native state end-to-end distance determined from the crystal structure. ΔL_{C_calc} is the calculated contour length change for each individual domain. The contour length changes for EF-G and *III-IV-V* are determined from force ramp experiments, displayed here as mean ± standard deviation.

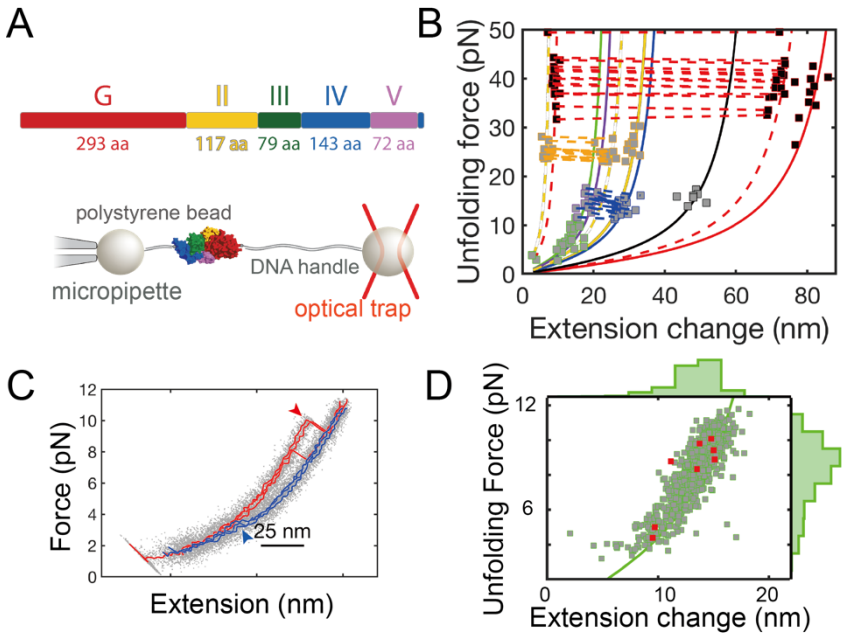


Figure VI-2 Domain III unfolding in full-length EF-G. (A) Single-molecule optical tweezers setup diagram (not to scale). The primary structure of EF-G is shown on top. (B) Scatter plot of initial transitions from full-length EF-G. Coloring for individual domains as in (A). The solid lines represent WLC models calculated for unfolding of individual domains (black: simultaneous unfolding of domains IV and V). The dashed WLC curves correspond to unfolding intermediates. Dashed lines connect transitions occurring in the same stretching cycle due to the population of an unfolding intermediate. (C) Example FECs from a single EF-G molecule with a force range of 2-12 pN, therefore only domain III unfolding (red) and refolding (blue) are observed. Grey dots are raw data (1000 Hz), while lines are filtered data (33 Hz). The red and blue arrowheads point to each one of the unfolding and refolding transitions respectively. (D) Scatter plot of domain III unfolding transitions. The red squares represent initial transitions, and the green squares are transitions after refolding. The force and extension change distributions are plotted on the right and on top.

VI-2.2: Domain III folding and unfolding follows a two-state model

Our previous work (Liu et al., 2017, 2019) showed that domain-wise cotranslational folding of the two N-terminal domains (G and II) is an important waypoint of EF-G folding. To determine whether domain-wise cotranslational folding continues through the remaining EF-G domains, domain III folding and unfolding were characterized. Repeated force ramp cycle between 2 and 12 pN were conducted, separated by a 0.5 s refolding pause at the minimum force. In this protocol, only domain III was selectively unfolded, while the remaining domains remained natively folded. Domain III refolded robustly, apparent as a “zip” in the relaxation

curves (blue arrowhead in Figure IV-2C). As a result, each pulling curve after refolding shows a characteristic unfolding transition for domain III (red traces in Figure IV-2C). Transitions after refolding (Figure IV-2D, green squares) were very similar to those obtained during initial unfolding (Figure IV-2D, red squares), suggesting that domain III efficiently refolds to its native state. No intermediate states during domain III unfolding and folding were detected. Domain III unfolding and folding therefore appears to be adequately described by a two-state model in which only the folded and unfolded states are populated.

VI-2.3: Domain III is mechanically unstable in G-II-III

Domain III forms extensive contacts with the other EF-G domains. The buried surface areas between domain III and the other domains were characterized using a tool developed by Krissinel et al. (Krissinel and Henrick, 2007) (Table VI-3). Given that the buried surface areas of domain III with the N-terminal domains are more than the C-terminal domains, we expected the N-terminal domains contribute more to domain III stability (Chen et al., 2013; Horton and Lewis, 1991).

		buried surface area ($Å^2$)
domain III	*G-II*	684
	IV-V	490

Table VI-3 Buried surface area calculation between domain III and the rest domains.

In order to determine how the N-terminal domains G and II affect domain III stability, we conducted experiments with a construct containing only the three N-terminal domains G, II, and

III (*G-II-III*, Figure VI-3A). Domain III folding and unfolding transitions in this construct occurred at very low forces. Instead of exhibiting well-separated folding and unfolding transitions in our non-equilibrium force ramp experiments, we observed hopping in the range of 2-6 pN. (Figure VI-3A bottom panel). Clear unfolding signatures of the G-domain and domain II confirmed that the molecule was properly tethered (Figure VI-3A and C). Some cycles did not show detectable transitions, suggesting that domain III has very low mechanical stability in *G-II-III*. This result suggests that domain III may not fold co-translationally. We hence conducted force ramp experiments with 531_{RNC}, a stalled ribosome-nascent complexes with the N-terminal *G-II-III* just outside the ribosomal exit tunnel (Figure VI-3B). Interestingly, we could not even detect transitions from domain III (Figure VI-3B). The transitions from domain II and the G-domain confirm that the N-terminal *G-II* folds to its native state in 531_{RNC}, similar to *G-II-III*. This observation is consistent with previous work indicating that the ribosome destabilizes the native state of domains that are in close proximity (Cabrita et al., 2016; Kaiser et al., 2011; Liu et al., 2017, 2019; Samelson et al., 2016).

Due to the low throughput of observing domain III folding and unfolding once we unfolded domains G and II, domain III was selectively unfolded by setting the pulling force range to 2 pN to 12 pN. The unfolding forces for these transitions in the *G-II-III* construct (with a mean force of only 3.8 pN) were indeed much lower than those observed in full-length EF-G (with a mean force of 8 pN) (Figure VI3-D). Taken together, these observations indicate that domain III stability is greatly reduced in the absence of domains IV and V, and is not likely to fold co-translationally.

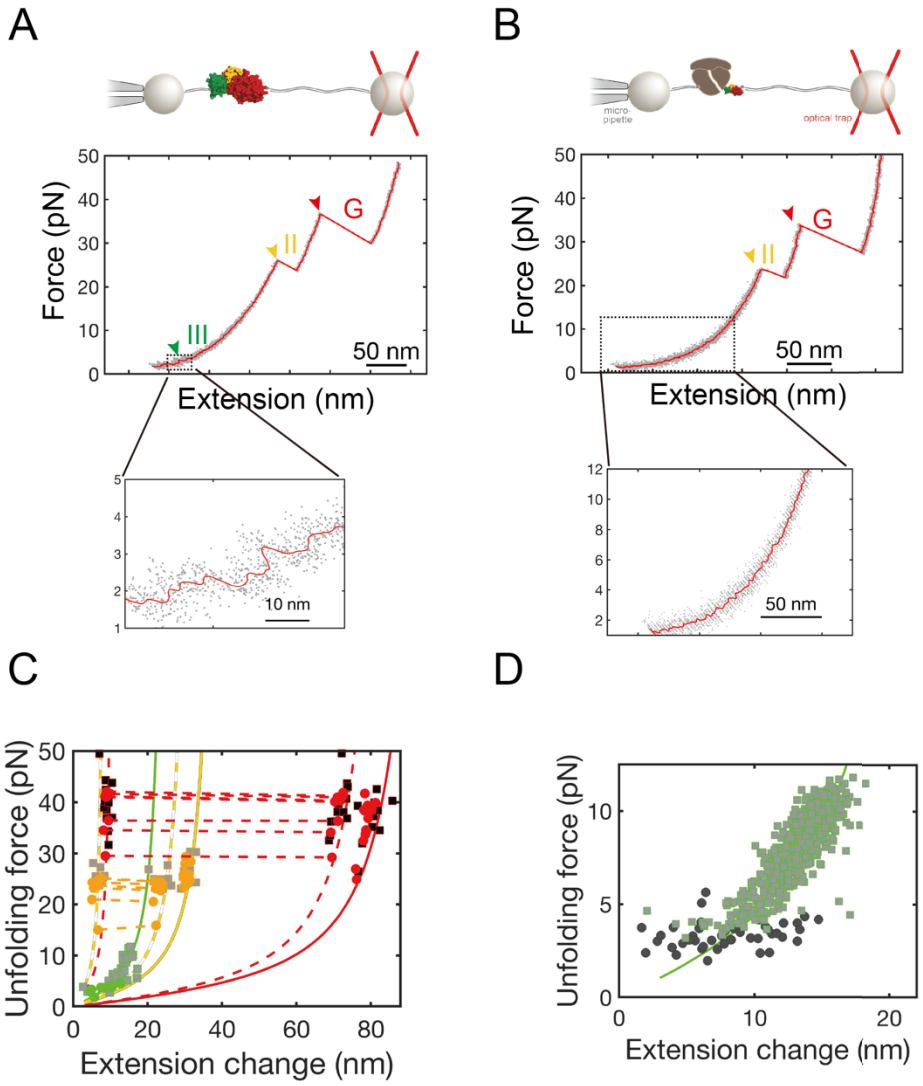


Figure VI-3 Domain III is less stable in *G-II-III* than in full-length EF-G. (A) Top panel: Experimental setup of *G-II-III* single-molecule optical tweezers experiment. Middle panel: Example FEC plot (stretching curve only) from the initial stretching of a single *G-II-III* molecule. Domain II and G-domain unfolding are apparent, but domain III unfolding occurs at very low force (inset). Grey dots are raw data (1000 Hz), red line is filtered data (33 Hz). Colored arrowheads indicate unfolding of each domain. (B) Top panel: Experimental setup of 531_{RNC} single-molecule optical tweezers experiment. Middle panel: Example FEC plot (stretching curve only) from the initial stretching of a single 531_{RNC} molecule. Domain II and G-domain unfolding are apparent, but domain III unfolding is missing (inset). Grey dots are raw data (1000 Hz), red line is filtered data (33 Hz). Colored arrows indicate unfolding of each domain. Coloring as in (A). (C) Scatter plot of *G-II-III* unfolding transitions (colored circles, same as the domain colors). They are overlapping with the G-domain, domain II and III unfolding transitions from full-length EF-G (colored squares), indicating that domain *G-II-III* folds to similar structure as EF-G. (D) Scatter plot of force vs extension change of domain III unfolding. Green squares represent transitions from full-length EF-G, black circles represent transitions from *G-II-III*. The unfolding force of domain III is much lower in *G-II-III* than full-length EF-G. The large variance of extension changes from domain III unfolding of *G-II-III* is partially due to higher noise in the lower force range.

VI-2.4: Domains in III-IV-V adopt their native structures

As shown in the previous section, *G-II* is insufficient to stabilize domain III. Domain III has even less amounts of buried surface areas with C-terminal domains IV and V. It was hence expected that *IV-V* has a similar low impact on domain III stability and that a combination of *G-II* and *IV-V* is necessary for stability. Surprisingly, experiments with a construct containing only domain III, IV and V (*III-IV-V*) revealed that all three domains in this construct were stably structured. *III-IV-V* unfolded at forces similar to those in full-length EF-G, both during initial stretching and after refolding (Figure VI-4A and B). Domain III refolded robustly after selective unfolding in the 2 to 12 pN range, similar to what was observed in full-length EF-G (Figure VI-4C and E). However, when all three domains in *III-IV-V* are mechanically unfolded, refolding was inefficient, even after a longer pause for refolding (5-10 seconds). In addition, the force extension curves after refolding do not always resemble the initial curves (Figure VI-4A), suggesting that the molecule populates non-native states. These experiments indicated that the domains in *III-IV-V* adopt the same native structures as in full-length EF-G (Figure VI-4B), that domain III mechanical stability is likely determined by the native interactions with folded domains IV and V, and that misfolding among the three C-terminal domains interferes with productive folding.

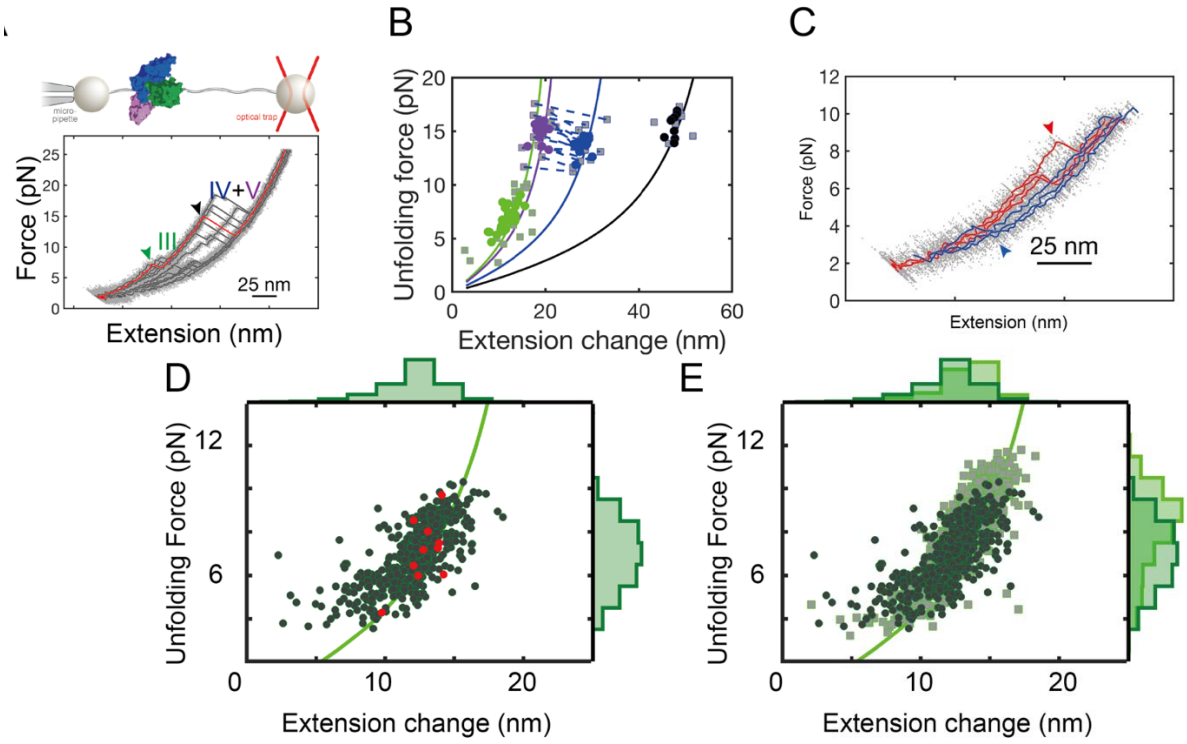


Figure VI-4 Domains in *III-IV-V* are natively folded. (A) Top panel: experiment setup of *III-IV-V* single-molecule optical tweezers experiment. Bottom panel, example FECs plot (stretching curve only) from a single *III-IV-V* molecule. The red curve represents the initial pull. The heterogeneous FECs indicate that *III-IV-V* refolds inefficiently and populates non-native states. Grey dots are raw data (1000 Hz), lines are filtered data (33 Hz). Colored arrows indicate unfolding of each domain in the initial stretching curve. (B) Scatter plot of *III-IV-V* unfolding transitions (colored circles, same as the domain colors). They are overlapping with the domain III, IV and V unfolding transitions from full-length EF-G, indicating that domain *III-IV-V* folds to similar structure as EF-G. (C) Example FECs from single *III-IV-V* molecule. In the force range of 2-10 pN, only domain III unfolding (red) and folding (blue) were observed. The red and blue arrows point to each one of the unfolding and refolding transitions respectively. (D) Scatter plot of domain III unfolding transitions from *III-IV-V*. The red circles are from initial stretching and the green circles are from transitions after refolding. The force and extension change distributions are plotted on the right and top. (E) Comparison of domain III unfolding between full-length EF-G (squares) and *III-IV-V* (circles). The force and extension change distributions are plotted on the right and on top (light green for full-length EF-G and dark green for *III-IV-V*).

VI-2.5: Domain III thermodynamic stability is derived from contacts with domains IV and V

Force ramp experiments are non-equilibrium measurements, as is apparent by the pronounced hysteresis between unfolding and refolding curves (Figure VI-2C, VI-4C). While these measurements readily yield kinetic rates, but do not directly provide thermodynamic

stabilities. To dissect the energetic contributions of native contacts to domain stability, the folding free energies must be determined. To overcome this limitation, we extended an approach described by Dudko and co-workers (Dudko et al., 2008) that relates the unfolding and refolding force distributions to force-dependent unfolding and refolding rates. The unfolding and refolding force distributions of domain III in either full-length EF-G or the *III-IV-V* construct were converted into force-dependent lifetimes (Figure VI-5). Fitting the lifetimes using Bell's model (Bell, 1978) (Figure VI-5) yielded the folding and unfolding rates at zero force (k_f^0, k_u^0) as well as transitions state distances (summarized in Table VI-4). The zero-force folding and unfolding rates allow us to determine the equilibrium constant ($K_{eq}^0 = \frac{k_f^0}{k_u^0}$), which directly relates to the thermodynamic stability. The thermodynamic stabilities of domain III in full-length EF-G and in the *III-IV-V* construct determined with this approach are very similar (-7.0±0.6 kcal/mol for EF-G, and -7.0±0.4 kcal/mol for *III-IV-V*). This observation indicates that the N-terminal *G-II* unit does not contribute much stability to domain III. The unfolding and refolding rates at zero-force of domain III in full-length EF-G and *III-IV-V* are indeed very similar as well (table VI-4), although native state to transition state distances differ slightly.

Constructs	k_u^0 (s^{-1})	k_f^0 (s^{-1})	$X_u^\dagger$ (nm)	$X_f^\dagger$ (nm)	ΔG_{cal} (kcal/mol)	ΔG_{CFT} (kcal/mol)
EF-G	0.019±0.016	3276±707	2.5±0.4	7.4±0.2	-7.0±0.6	~-6.6
III-IV-V	0.021±0.007	2893±1043	3.0±0.2	7.2±0.4	-7.0±0.4	~-5.9

Table VI-4 Unfolding and folding kinetics of domain III in EF-G and *III-IV-V*.

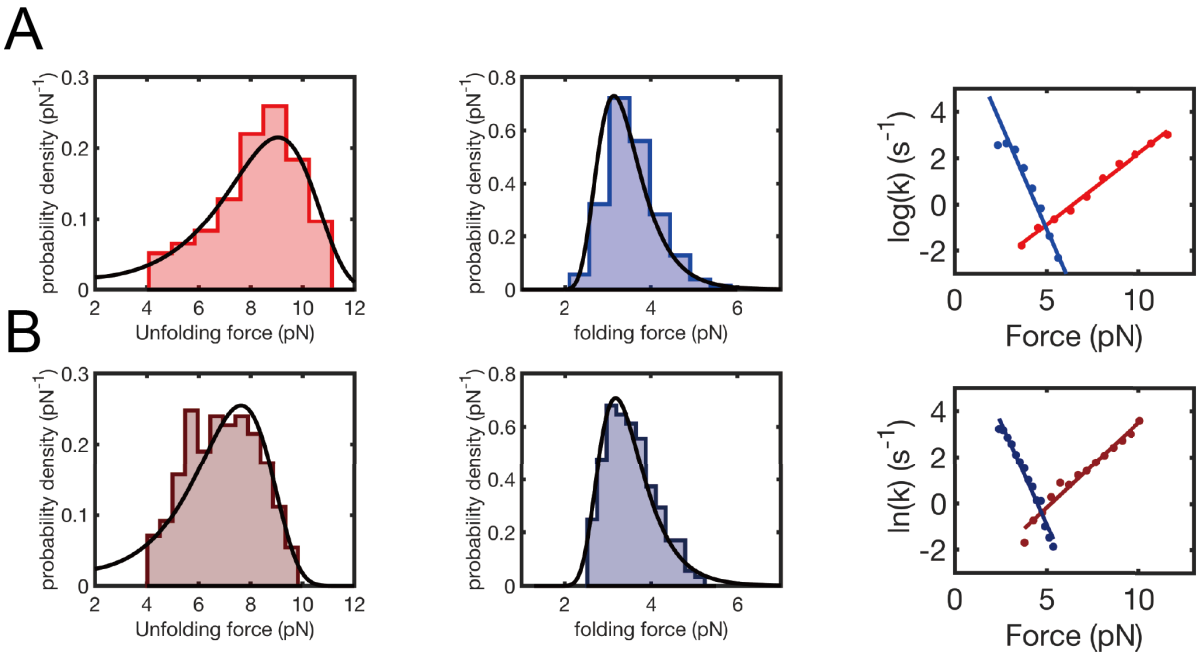


Figure VI-5 Estimation of domain III unfolding and folding rates. (A) Analysis on EF-G unfolding and refolding force distributions. Left: domain III unfolding force distribution of EF-G (red). The black curve is derived from Dudko's method. Middle: domain III refolding force distribution of EF-G (blue). The black curve is the fit using Dudko's method. Right: chevron plot of force-dependent unfolding (red dots) and refolding rates (blue dots) of domain III in full-length EF-G. The data are fit with Bell's model with red line for unfolding and blue line for refolding. (B) Analysis on *III-IV-V* unfolding and refolding force distributions. Panels as in A.

To complement the analysis based on Dudko's transformation, constant force measurements were conducted where domain III folding and unfolding transitions were monitored while keeping *IV-V* intact (Figure VI-6A). Using a non-parametric Bayesian Hidden Markov Model (Sgouralis et al., 2018), the rates at two forces (4.5 and 5 pN) were extracted

from these constant force trajectories. The rates determined by the two approaches were in good agreement (Figure VI-6B, circles vs. triangles). Thus, the direct measurement of kinetic rates at two different forces validates the Dudko's approach, which spans a larger force range.

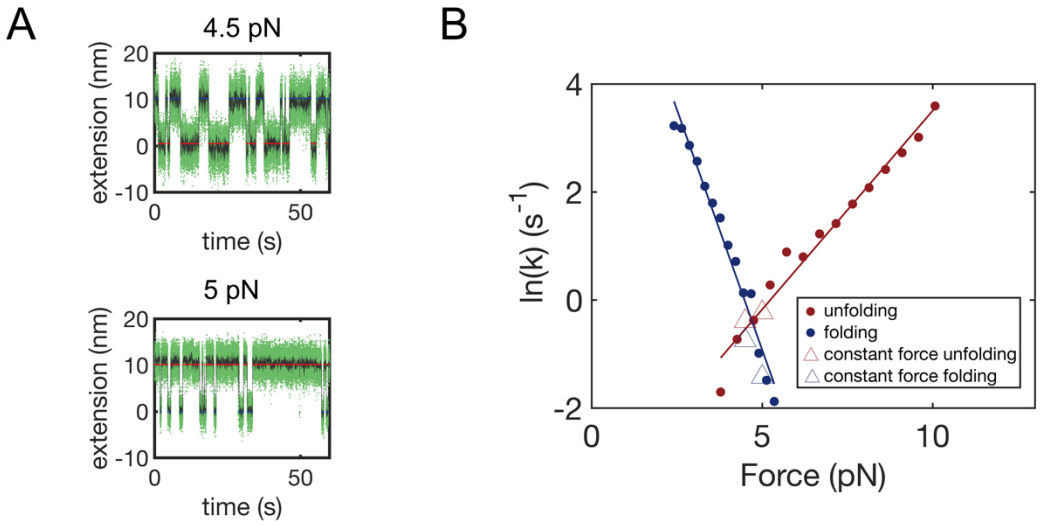


Figure VI-6 Domain III folding and unfolding rates estimation. (A) Example folding trajectories of domain III at 4.5 pN (top panel) and 5 pN (bottom panel) in *III-IV-V*. Green dots are raw data with sampling rate of 1000 Hz, black line is filtered data with 33 Hz. The blue and red lines are assigned unfolded state (blue) and folded state (red). (B) ln(k) vs force plot of domain III unfolding (red) and folding (blue) from *III-IV-V*. The dots are from the analysis of force distributions using Dudko's method, lines are based on linear fit using Bell's model (same as in Figure VI-5B).

To corroborate these findings, an alternative approach was utilized to extract the thermodynamic stability of domain III from the force ramp data. The Crooks fluctuation theorem relates the distributions of irreversible work during mechanical unfolding and refolding to the equilibrium free energy of the underlying process (Collin et al., 2005; Crooks, 1999; Jarzynski, 1997; Liphardt, 2007; Shank et al., 2010) (Figure VI-7). With this approach, the equilibrium free energy at zero-force was estimated by determining where the unfolding and refolding work distributions intersect using linear interpolation (Figure VI-7). The free energy values we obtained (-6.6 kcal/mol for EF-G, and -5.9 kcal/mol for *III-IV-V*) are consistent with those estimated by the extrapolation method using Bell's model (see above). Overall, we concluded

that domain III has similar stabilities in both full-length EF-G and *III-IV-V*, which is about -6 to -7 kcal/mol. Due to limited statistics at the crossing region of the unfolding and refolding work distributions, the error of the free energy could not be estimated accurately.

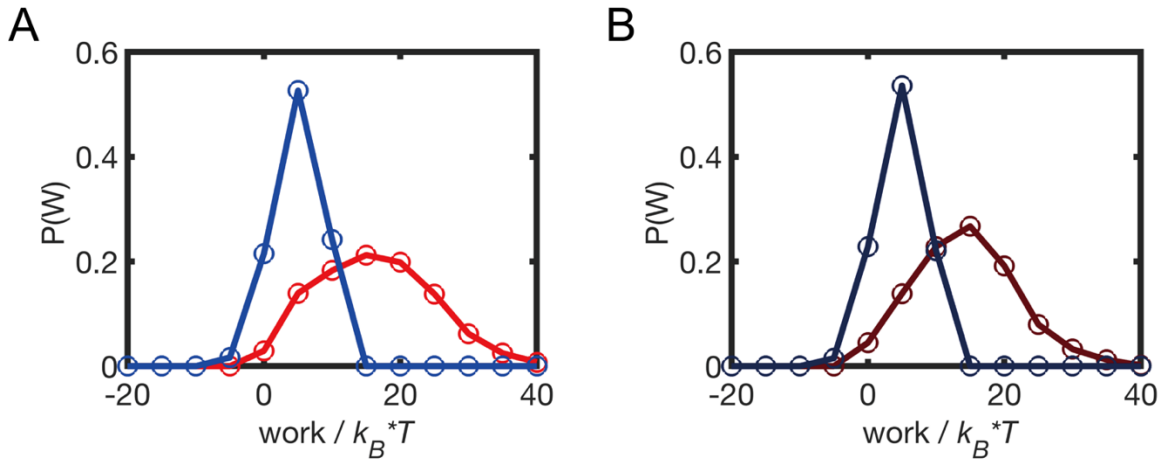


Figure VI-7 Estimation of domain III stability using Crooks fluctuation theorem. (A) Domain III unfolding work distribution (red) and refolding work distribution (blue) in full-length EF-G. (B) Domain III unfolding work distribution (red) and refolding work distribution (blue) in *III-IV-V*. The experiments were done at room temperature (298 K).

VI-2.6: Unfolded neighboring domains impair domain III folding

How the structural state of a domain influences folding of a neighboring domain is not well understood. Previous work with multidomain proteins demonstrated that unfolded neighboring domains can either not affect (Batey and Clarke, 2008), decelerate (Jahn et al., 2016, 2018; Liu et al., 2019) or accelerate domain folding (Randles et al., 2008). To dissect the interactions of domains in the C-terminal part of EF-G, we compared how the state of domains IV and V (folded vs. unfolded) affects domain III folding in full-length EF-G and *III-IV-V*. When domains IV and V had been mechanically unfolded (fully unfolded state) before refolding was initiated, a set of transitions was observed that did not match those of any of those representing unfolding of the native domains were observed (Figure VI-8A and B compare to Figure VI-4B).

The fact that these transitions occurred at low force suggested that stabilizing contacts from domain V and IV were missing (Figure VI-8A and B). This observation also implied that misfolding occurs when all the three domains were unfolded. Consistent with this hypothesis, stable folding of *III-IV-V* from fully unfolded state within 60 seconds at 3 pN was observed rarely (Figure VI-8C). The observed intermediate states may be the same states that unfolded at low force in force ramp experiments (open circles in Figure VI-8A and B).

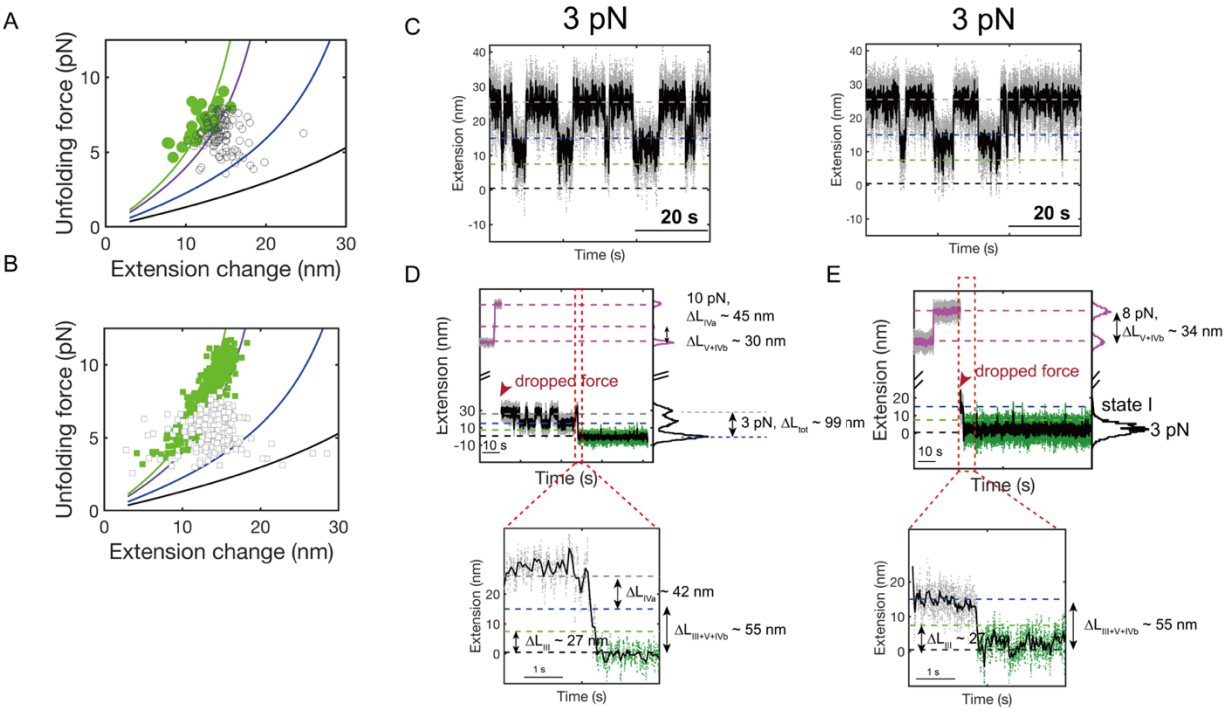


Figure VI-8 Domain III cannot stably fold before folding of *IV-V*. Comparison of domain III unfolding transitions with domains IV and V unfolded (open squares and circles) or folded (colored squares and circles) for *III-IV-V* (A) and EF-G (B). For both constructs, the contour length changes are longer and unfolding forces are lower when domain *IV-V* is unfolded for both constructs. Some of the transitions show contour length changes that are consistent with native unfolding of domain V+IVb (domain V with IVb) and IVa, but occur at much lower force. (C) Two examples showing hopping of *III-IV-V* at 3 pN after complete unfolding of all three domains. The starting point is the fully unfolded state where all three domains are unfolded. The dashed lines correspond to the fully unfolded state (grey), IVa folded state (blue), both domains IV and V are folded but domain III is unfolded (green), and all domains fully folded state (black). Grey dots: taw data (1000 Hz), black lines: filtered data (33 Hz). Most often, *III-IV-V* does not fold within 60 seconds. (D) Unfolding of all domains of *III-IV-V* at 10 pN, followed by refolding of *III-IV-V* when fully unfolded *III-IV-V* was held at constant force of 3 pN. Dots: raw data (1000 Hz), solid lines: filtered data (33 Hz). Calculated extensions for different states are indicated by the dashed lines with the same colors as in (C). The total contour length change from highest extension (fully unfolded) to the lowest extension (folded) is about 100 nm, as expected for the construct *III-IV-V*. Zoomed in trajectory (bottom panel)

showing folding at the order of first domain IVa, then domains V with IVb and domain III. Individual step sizes are as expected. In this example, at least two intermediate states populated. The raw data where domain III stably folded was plotted as green dots. (E) Unfolding of only domain V+IVb at 8 pN with domain III already being unfolded before holding the force at 8 pN, followed by dropping the force to 3 pN to monitor the folding. The unfolding at 8 pN matches well with unfolding of domain V+IVb. The dashed lines and coloring scheme are the same as in D The state I is indicated by the blue dashed line. The filtered data of unfolding of V+IVb at 8 pN is plotted as a pink line.

In a few cases where stable folding of *III-IV-V* from fully unfolded state was observed, the major "intermediate state" that was populated prior to folding had an extension (left panel of Figure VI-8D) very similar to that of a state in which only domain IVa is folded (state I, with folded IVa but unfolded domain III and V+IVb, indicated by blue dashed line in the right panel of Figure VI-8D). To determine whether this state is a misfolded state or the intermediate state I, which is a productive on-pathway intermediate, we conducted constant force measurement that started from the state I. To this end, we allowed the molecule to unfold only partially at the unfolding force, and then immediately switched to the refolding force (Figure VI-8C right panel). Under these conditions, where folding initiates from state I, we always observed the complete folding of *III-IV-V* within a short period of time. By performing a one-sided binomial test with the null hypothesis of that the major intermediate state we observed when folding started from the fully unfolded state is the same as state I, we obtained a p-value of 1×10^{-14}, suggesting that *III-IV-V* very likely misfolds when folding begins from the fully unfolded state. After folding, domain III still fluctuates between the unfolded state and the folded state after folding of *III-IV-V*, as is as expected at this force (Figure VI-8D, compared to Figure VI-10). Taken together, domain III can only stably fold after domains IV and V are folded.

VI-3: Discussion

Here, folding events during the late stages of EF-G synthesis have been dissected here, revealing an intricate coupling between domains III, IV and V that make up the C-terminal part

of the molecule. Single-molecule experiments revealed that domain III stability depends on extensive contacts with a unit comprised of domains IV and V, which themselves are energetically coupled due to their primary structure topology. Energetic coupling previously observed in the N-terminal half of EF-G (G-domain and domain II, see chapter V). In both cases, a large amount of buried surface area indicates extensive contacts between the folded domains. The tight coupling between these two pairs of domains agree with the notion that *G-II* and *III-IV-V* are forming super-domains (Chen et al., 2013; Lin et al., 2015). For *G-II*, the domains fold in the order in which they emerge from the ribosome. In contrast, the dependencies in the C-terminal part are such that the order of folding does not match order of synthesis. Domain III can fold only after synthesis of domains IV and V. As a result, folding of the C-terminal half of EF-G has to be completed post-translationally.

Our approach of manipulating single molecules allows us to exploit the differential mechanical stabilities of individual domains and unfold them selectively while keeping the rest of the molecule folded. I used this approach to characterize the folding of domain III in detail. Using two independent approaches, I found that domain III has similar thermodynamic stabilities regardless of whether or not the N-terminal *G-II* super-domain is present. However, the stability is much reduced in the absence of the C-terminal *IV-V* unit. Interestingly, these findings demonstrate that *G-II* makes only small contributions to domain III stability, even though it forms an extensive interface with domain III in the native EF-G structure.

A comparison of full-length EF-G and *III-IV-V* revealed differences in the transition state distances for domain III unfolding, but similar zero-force unfolding rates (Table VI-4). The apparent differences in transition state distances might be due to slightly different pulling axes in

the full-length EF-G and *III-IV-V* constructs (Brockwell et al., 2003). The observation that domain III folds at similar rates in full-length EF-G and *III-IV-V* is different from what has been observed on spectrin (Batey and Clarke, 2008; Randles et al., 2008), which found that the native interactions from the N-terminal domains accelerates folding of the C-terminal domain. However, our results are consistent with the concept of minimal frustration theory (Bryngelson and Wolynes, 1987), which implies that the folding rates of natural domains are optimized through evolutionary selection, and the thermodynamic stabilities of natural proteins are primarily determined by unfolding rates (Tzul et al., 2017).

Consistent with the above conclusion that the two N-terminal domains G and II do not contribute significantly to domain III stability, we did not detect any folding or unfolding transitions of domain III in the 531_{RNC} nascent chain, even though domains G and II fold to their native states. In addition, we also observed much lower domain III unfolding force in *G-II-III* comparing to full-length EF-G (Figure VI-9), which suggests a higher unfolding rate of domain III in *G-II-III*. Domain III in *G-II-III* exhibits hopping behavior in the force range where it is stably folded when domains IV and V are present. This observation indicates that domain III is competent to fold in *G-II-III* even in the presence of mechanical load (Figure VI-3). Consequently, domain III stability appears to rely on interactions with folded domains IV and V.

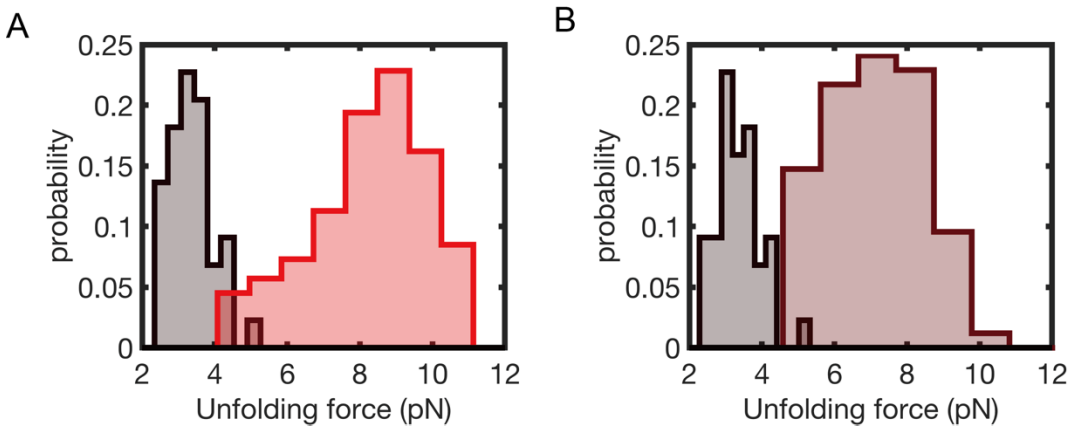


Figure VI-9 Domain III unfolds at lower force in *G-II-III*. (A) Comparison of domain III unfolding force distributions for *G-II-III* (black) and full-length EF-G (red). (B) Same as (A) for *G-II-III* (black) and *III-IV-V* (dark red).

Constant force measurements conducted at 3 pN also support the pronounced effect of these C-terminal units on domain III stability (Figure VI-10). In the context of cotranslational folding, this energetic dependency implies that folding of domain III can occur only post-translationally, because it relies on the critical interface with the C-terminal *IV-V* unit, which is synthesized last.

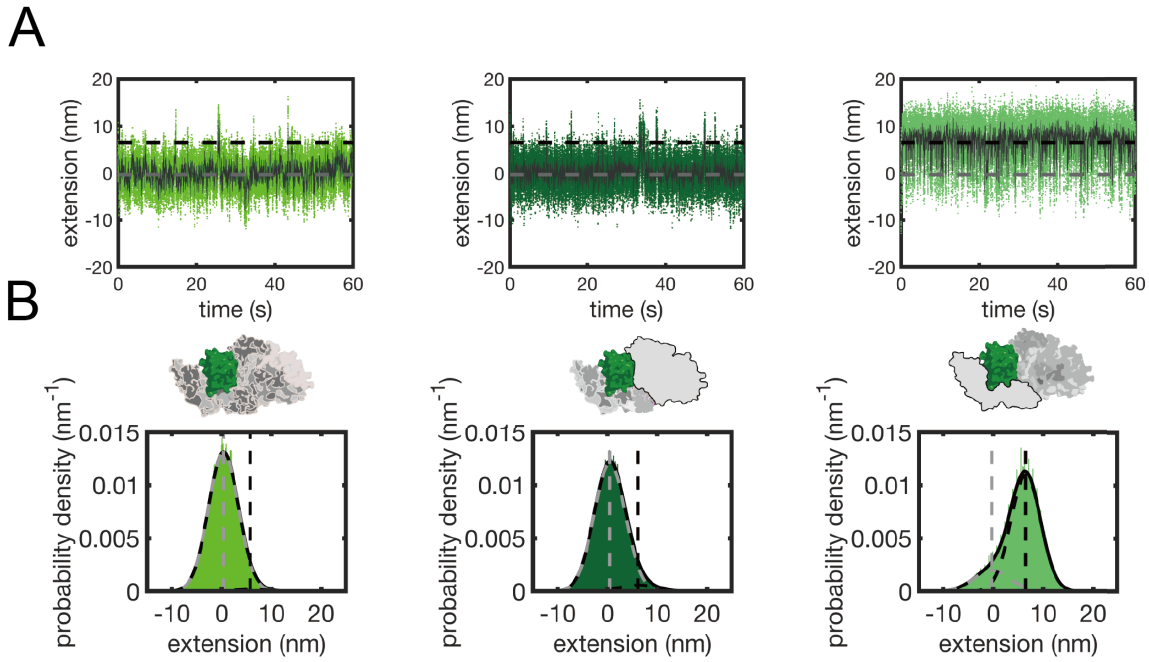


Figure VI-10 Domain III folding equilibrium at 3 pN. (A) Example traces of domain III dynamics at 3 pN from three different constructs. Left: full-length EF-G, middle: *III-IV-V*, right: *G-II-III*. (B) Extension distributions for traces shown in (A). The structure diagrams of each construct are plotted on top of the diagrams, with domain III colored in green, and the other domains colored in grey. The more stable state of domain III is the folded state in full-length EF-G (left) and *III-IV-V* (middle), but the equilibrium is shifted towards unfolded state in *G-II-III* (right).

Finally, we observed that complete refolding of the *III-IV-V* construct is not efficient (Figure VI-4A). During the refolding pause when folding started from the fully unfolded state, the molecule mostly populated states that were non-native and rapidly unfolded at low forces (white squares in Figure VI-8A and open circles in Figure VI-8B). One possible reason for this observation is inter-domain misfolding, which has been observed to slow down productive folding in a number of multidomain proteins (Borgia et al., 2011; Jahn et al., 2016; Kumar and Chaudhuri, 2018; Liu et al., 2019; Zheng et al., 2013). The observation that folding at 3 pN within 60 seconds has a probability of only around 0.1, i.e. folding is observed in only six out of sixty attempts under these conditions, also supports the conclusion that the inter-domain misfolding occurs when *III-IV-V* folding begins from the fully unfolded state. Directly monitoring the folding trajectories where folding initiated from the state I of *III-IV-V* (Figure VI-8D right panel) allowed us to conclude that stable folding of domain III is likely to occur after folding of domain IVa.

In conclusion, our studies provide an example of asymmetric energetic contributions from neighboring domains to the stability of individual structural units in multidomain protein. Domain III is unstable when folded *IV-V* is missing. This observation is consistent with a postulated evolutionary rule that domains incapable of independent folding are stabilized by favorable interactions with neighboring domains (Bhaskara and Srinivasan, 2011). EF-G undergoes large conformational changes between two super-domains *G-II* and *III-IV-V* during translation elongation (Lin et al., 2015). The dynamic behavior of domain III may be necessary for achieving this conformational flexibility (Chen et al., 2016). At the same time, this requirement complicates folding to the native state, because it precludes domain-wise co-

translational folding beyond the *G-II* unit. Perhaps the observed properties reflect a compromise between functional and folding constraints, with the latter being alleviated by interactions with the ribosome and molecular chaperones.

VI-4: Materials and methods

VI-4.1: Unfolding transition analysis (force ramp experiments)

The transitions are determined as reported in Liu *et al.* 2019 (Liu et al., 2019). In brief, during force ramp experiments, we continuously move the optical trap at constant velocity (100 nm/s). Unfolding transitions are apparent as "rips" in the force extension curves (FECs). Each rip is characterized by an unfolding force and a change in molecular extension (Figure VI-11). We determined the F_{unf} by locating discontinuities in the FECs using a custom Matlab script. To determine the Δx_{unf}, we first locally fit the data before and after rip (Figure VI-11) with at least 5 data points for each state, then we measured the difference in extensions before and after rip at F_{unf} (Figure VI-11). The rips were selected manually, which were apparent in discontinuity of the force (with a criterion of $\Delta F > 0.5$ pN for adjacent points that are before and after rips).

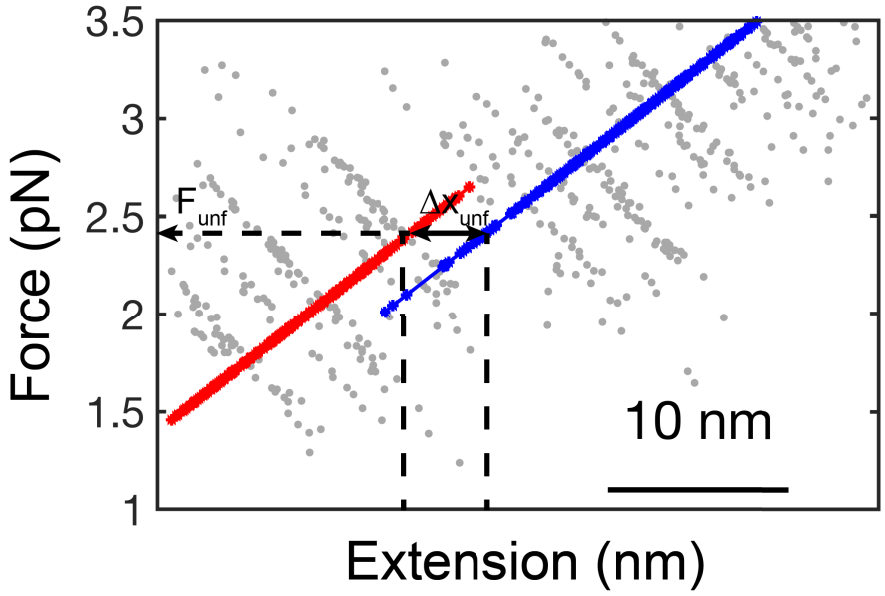


Figure VI-11 Unfolding transition determination. Grey dots are raw data with sampling rate of 1000 Hz. Blue and red lines are linear fit to the corresponding states. The unfolding force and extension change are determined as described in the text.

VI-4.2: *Estimation of domain III stability using Crooks fluctuation theorem*

We adopted the alternate bounds of integration for the Crooks fluctuation theorem method as described by Shank, E., et al (Shank et al., 2010). Briefly, the raw traces were fitted by lines around the rips' locations. The force where integration was done was determined by the rip force. The integration starts at the rip and ends at the other state at the same force. Since the force at the boundaries were the same, the work went into the system equals to unfolding/refolding of the protein plus the work needs to stretch/relax the same unfolded polypeptide to the rip force. The work needs to stretch/relax the unfolded polypeptide was estimated based on worm-like chain (WLC) model (Bustamante et al., 1994).

VI-4.3: *Folding rate estimation from constant force measurement*

We adopted the xl-ICON (Sgouralis et al., 2018), which can help us determine different states from each single molecule constant force folding trajectories using nonparametric Bayesian analysis. We then aggregated the lifetimes from the same forces for each state to transit to another specific state and fit with single exponential curves. The mean value will the folding average dwell time, which is the inverse of the corresponding transitions rates.

Chapter VII. Conclusions

VII-1: Summary

I have successfully developed EF-G as a new biophysical model for multi-domain protein folding studies. I found that folding of the N-terminal G-domain is likely a key folding waypoint for EF-G folding. The two N-terminal domains G and II are likely to fold in a domain-wise fashion. If co-translational folding fails, misfolding severely reduces the folding rate, which may have profound effect on EF-G productive folding. Interestingly, the ribosome modulates the folding of nascent EF-G in a length dependent manner. When the G-domain is in close proximity to the ribosome, the ribosome slows down its folding. This is consistent with the current understanding of that the ribosome favors the unfolded state of nascent chains in general. However, the underlying mechanism is still elusive. When it is farther away, the intrinsic folding rate is recovered. In the case of EF-G, the ribosome also suppresses misfolding between the first two domains, thus accelerating the folding of G-domain comparing to the isolated *G-II*. The mechanism for the suppression of misfolding might be the same as that for decelerating folding. Surprisingly, the folding of EF-G is not unidirectional. Emerging domain II can denature the already folded G-domain. This denaturation is not suppressed by the ribosome, while the chaperone TF blocks this denaturation. Therefore, we propose that an important role for nascent chain binding chaperones is to protect already folded structures. On the contrary, the C-terminal domains of EF-G are likely to fold post-translationally, because the inter-domain interactions between domain III and the C-terminal domains are critical for domain III stability. Therefore, the folding mechanism is determined by domain-domain interactions (Figure VII-1).

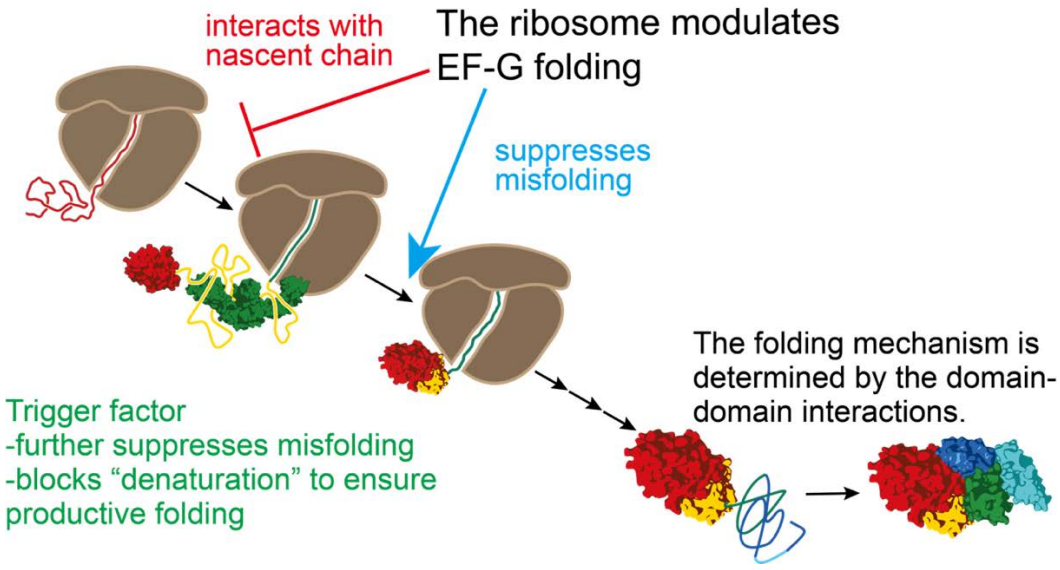


Figure VII-1 Conclusion model for EF-G folding. In brief, when the full sequence of the G-domain comes out of the ribosomal exit tunnel, folding is delayed by the interactions between nascent G-domain and the ribosome. However, this delay is alleviated when G-domain is farther away from the ribosome, therefore G-domain is likely to fold before full domain II sequence comes out of the tunnel. The presence of the ribosome and TF protects the folded G-domain and suppresses the misfolding during the process to ensure productive folding. However, domain III cannot stably fold before domains IV and V come out of the tunnel. More chaperones are likely to be required during the elongation process after domain III sequence comes out of the tunnel.

VII-2: Broader implications

Before my work, multidomain protein folding studies by biophysical techniques have been focusing mainly on small proteins in isolation. I have successfully developed EF-G as a model protein for multidomain protein folding studies in the context of the ribosome and a molecular chaperone. The most surprising finding is that an already folded domain can be denatured by the emerging sequence from the subsequent domain. It will be interesting to investigate in future studies how cells solve this problem and maintain protein homeostasis from cell biology perspective. Moreover, I found that the TF suppresses denaturation, which reveals a new role for nascent chain binding chaperones. Previously, the main focus of studies on molecular chaperones was to investigate how they increase the folding efficiency of their clients, and how they prevent

or reverse misfolding and aggregation. However, the aspect of protecting the already folded proteome was missing.

I also discovered that how the ribosome modulates folding of multidomain protein depends on nascent chain length. The ribosome accelerates the folding of a distal domain by suppressing the inter-domain misfolding. The fine-tuned modulation of nascent chain folding by the ribosome has been reported by other groups as well using different systems (Holtkamp et al., 2015; Kim et al., 2015). Similarly, studies have also focused on how tRNA abundance (Zhang et al., 2009) and synonymous codon (Buhr et al., 2016; Sander et al., 2014) affect the productive folding . This leads to the question of whether the tuning of the nascent chain folding by the ribosome is due to an evolutionary selection. A recent genome-wide study has supported this idea (Jacobs and Shakhnovich, 2017).

Finally, I have been able to demonstrate that the folding mechanism of a multidomain protein is determined by the domain-domain interactions, which might represent a general rule. The two N-terminal domains G and II are likely to fold co-translationally in a domain-wise manner. Domain II by itself is not very stable, but the correctly formed interface from the G-domain helps domain II reach the final stably folded state. In contrast, domain III stability derives from its interactions with C-terminal domains IV and V, thus domain III is likely to fold post-translationally. A similar conclusion has been drawn from a study on eEF-2, a homolog of EF-G in eukaryotes (Mönkemeyer et al., 2019). It will be interesting to look at other EF-G homologs to see if this rule is conserved. Domain III is highly dynamic supported by the fast exchange between its folded and unfolded states at low force. Another biophysical study using single-molecule polarized total internal reflection fluorescence (polTIRF) microscopy suggested that domain III is the most

dynamic domain in EF-G (Chen et al., 2016), which is likely relevant to the function of EF-G. A recent structural study has reported that EF-G have very distinct conformations in the pre- and post-translocation ribosome complexes, with the former one being much more compact (Lin et al., 2015). It will be interesting to see if the dynamic nature of domain III is essential for the relatively large rotation between the N-terminal *G-II* unit and the C-terminal domains of EF-G.

VII-3: Future directions

One of the remaining puzzles I would like to solve is to determine the molecular mechanism of the domain denaturation. Is it sequence dependent or length dependent or a combination of the two? Experiments with different sequences or sequences with different lengths replacing domain II sequence may help us to address this question. Can we better quantify the rate of the denaturation? How general is the denaturation in the proteome? Once we have a better idea of the molecular mechanism of denaturation, bioinformatics studies will be helpful to identify the generality of denaturation.

To Better characterize the interactions between the ribosome and nascent chains is also a key goal in the field, which requires a combination of multiple techniques. For instance, combining single-molecule force spectroscopy and fluorescence spectroscopy together will help us monitor interactions with a better resolution at atomic level. In addition, complementary ensemble methods such as NMR spectroscopy will provide atomic level information that single-molecule force spectroscopy cannot reach currently. Real-time experiments monitoring EF-G folding using more advanced instrumentation will overcome the current limitations of relying on stalled ribosome-nascent complexes, and help to define key folding events of multidomain proteins more accurately.

In addition, the molecular mechanism of how TF helps folding of its clients is still unclear. I found that TF may restrict *G-II* in its binding pocket to promote folding. Is this restriction related to the apparent reduction of misfolding between the two domains? Or does it result in other effects such as reducing the conformational search space? How exactly TF and EF-G interact is still unclear, and requires further biochemical and biophysical measurements to characterize. Experiments with different TF mutants that lack functions of individual domains of TF will help us identify the roles of

these domains in promoting the folding of EF-G. In addition, it is very interesting to expand similar studies to other nascent chain binding chaperones, such as DnaK, to investigate how do they affect the folding of EF-G. Finally, experiments with homologs of EF-G will allow us to draw more general conclusions on multidomain protein folding.

Appendices

A-1 Primer sequences and oligonucleotides

A-1.1 Primers for Gibson assembly

oligonucleotides	sequence	usage notes
Backbone_fwd	5'- ACGATATCAAGTGGTGGCTCCGGCG -3'	For Backbone PCR
Backbone_rv	5'- CGCGGCCGCGGATCCACC -3'	For Backbone PCR
EF-G_fwd	5'-ggatccgcggccgcgATGGCTCGTACAACACCC-3'	For EF-G insert PCR
EF-G_rv	5'-agccaccacttgatatcgtTTTACCACGGGCTTCAATTAC-3'	For EF-G insert PCR
G_fwd	5'-ggatccgcggccgcgATGGCTCGTACAACACCCATC-3'	For G-domain insert PCR
G_rv	5'-agccaccacttgatatcgtCGCAGGTACGTCAACCGG-3'	For G-domain insert PCR
G-II_fwd	5'-ggatccgcggccgcgATGGCTCGTACAACACCCATC-3'	For G-II insert PCR
G-II_rv	5'-agccaccacttgatatcgtTTCCATACGTTCCAGAATG-3'	For G-II insert PCR
G-II-III_fwd	5'-ggatccgcggccgcgATGGCTCGTACAACACCCATC-3'	For G-II-III insert PCR
G-II-III_rv	5'-agccaccacttgatatcgtCTGCGGTTTACCTACGTTC-3'	For G-II-III insert PCR
II_fwd	5'-ggatccgcggccgcgCTGCCATCCCCGGTTGAC-3'	For domain II insert PCR
II_rv	5'-agccaccacttgatatcgtTTTACCACGGGCTTCAATTAC-3'	For domain II insert PCR

Table-A 1 Primers for Gibson assembly.

A-1.2 Primers for RNCs template pcr

Oligonucleotide name	sequences
T4Ls_T7_fw	5'-GTCGGCGATATAGGCGCCAG-3'

328RNC_rev	5'-CGGGTCGGTAGCGATTTTGAACG-3'
358RNC_rev	5'-CTCACGTGCAGCTTTCACGGA-3'
386RNC_rev	5'-GATAGCAGCAGCGATGTCGC-3'
415RNC_rev	5'-TACCGGCTCAGGGAATTCCATACG-3'
452RNC_rev	5'-GATTCTTCGTCAGTCCATACACGG-3'

Table-A 2 Primers for RNCs template PCR.

A-1.3 Oligonucleotides for DNA handle generation

oligonucleotides	sequence	usage notes
NH-fw3	5'-/5AmMC6/CACCGAAACGCGCGAGGCAGCAG -3'	For cross-linking to CoA
cNH-fw3	5'-/5Phos/GTCGCTGCTGCCTCGCGCGTTTCGGTG -3'	For annealing to NH-fw3
NH-fw2	5'-/5AmMC6/CTGCTGGGGCAAACCAGCGTGGAC-3'	For cross-linking to polystyrene beads
cNH-fw2	5'-/5Phos/CGACGTCCACGCTGGTTTGCCCCAGCAG-3'	For annealing to NH-fw2
hfw2	5'-/5Phos/CTGCTGGGGCAAACCAGCGTGGAC-3'	For long DNA handle bead pcr
bio-rev2	5'-/5bio/CACCTGACCCCATGCCGAACTC-3'	For long DNA handle bead pcr

Table-A 3 Oligonucleotides for DNA handle generation.

A-2 Plasmids maps

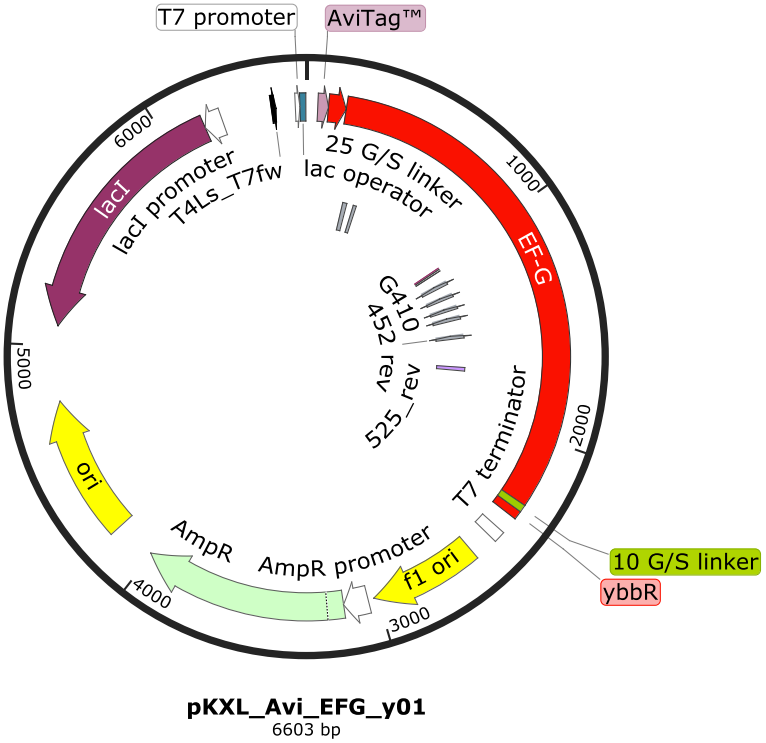


Figure-A1 Plasmid map for RNCs PCR templates.

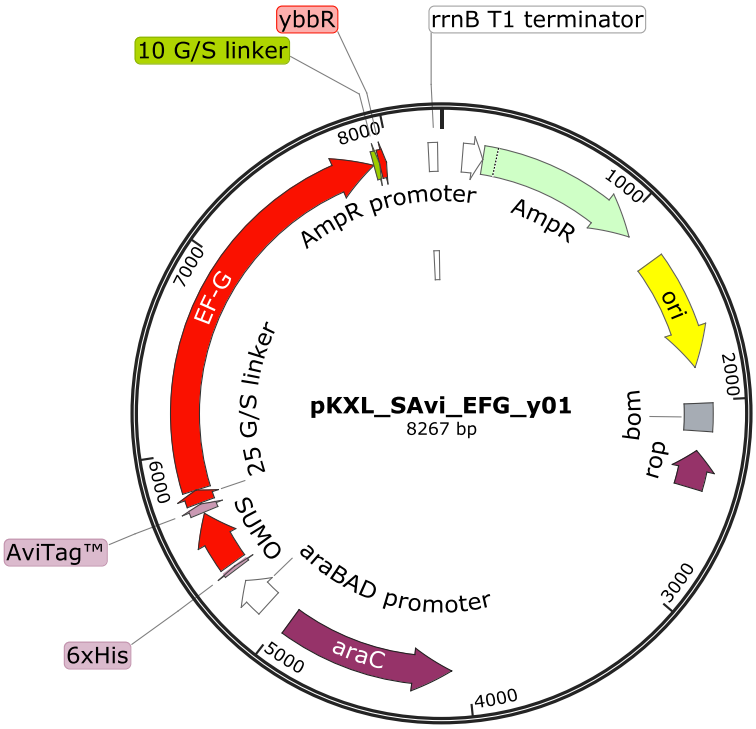


Figure-A2 Plasmid map for full-length EF-G.

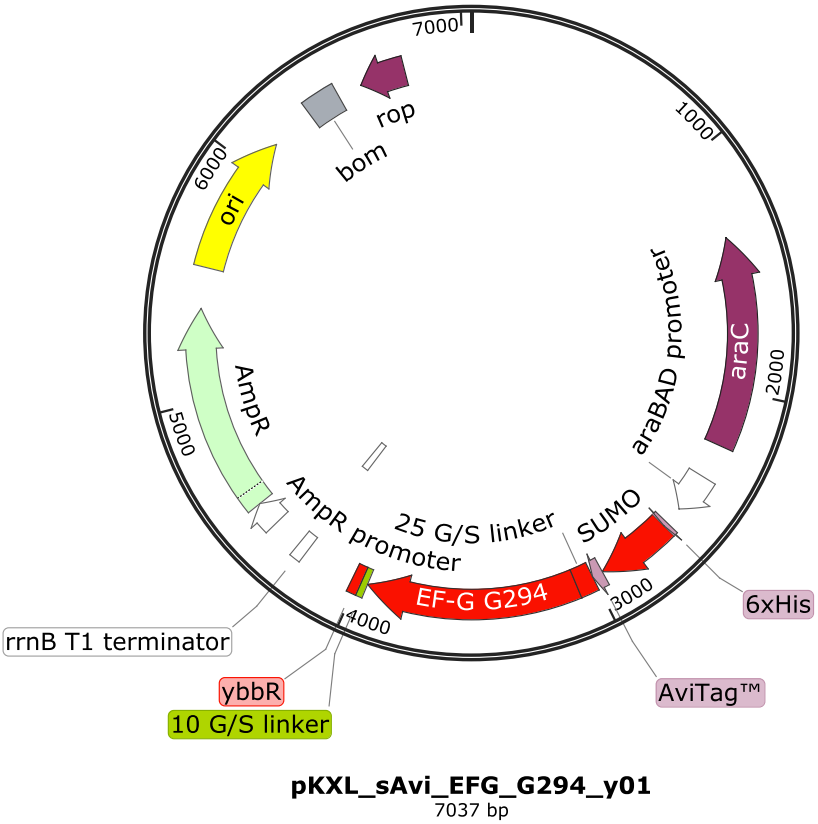


Figure-A3 Plasmid map for G-domain.

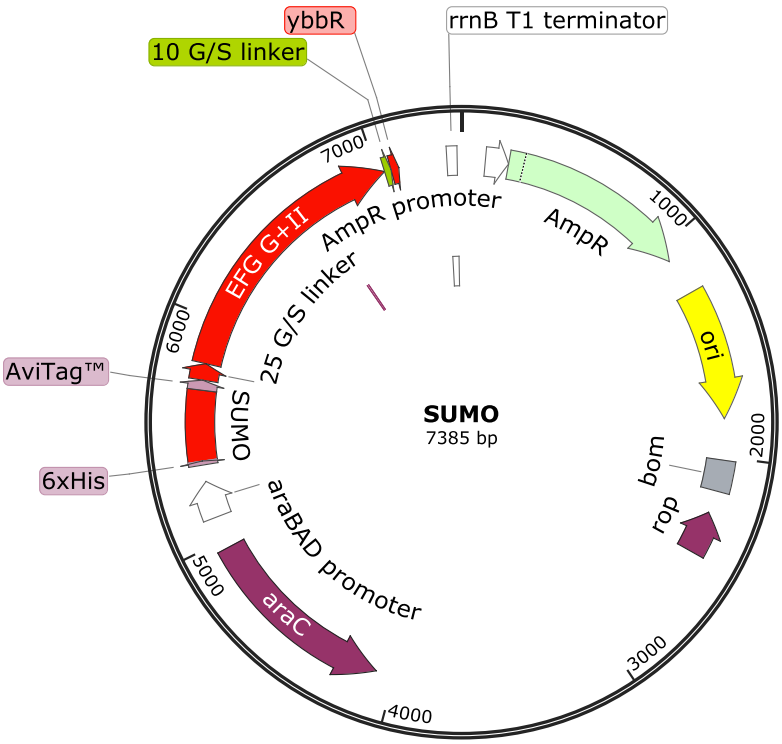


Figure-A4 Plasmid map for *G-II*.

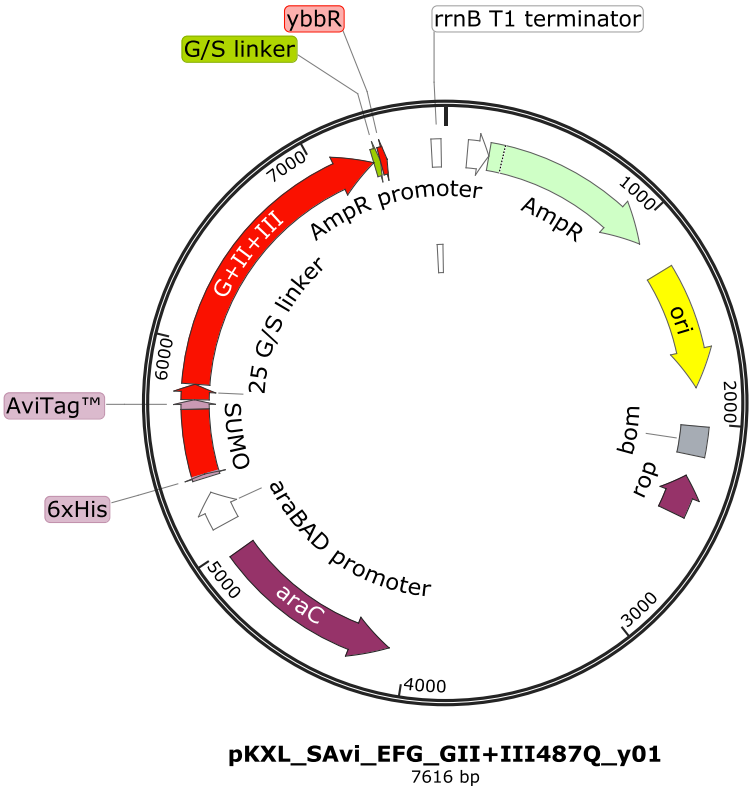


Figure-A5 Plasmid map for *G-II-III*.

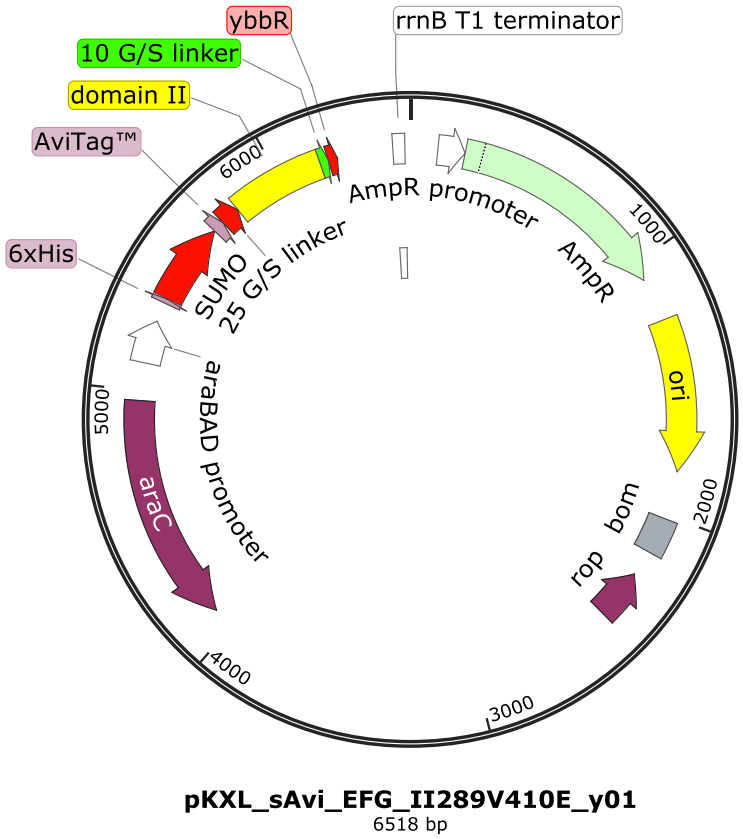


Figure-A6 Plasmid map for domain II.

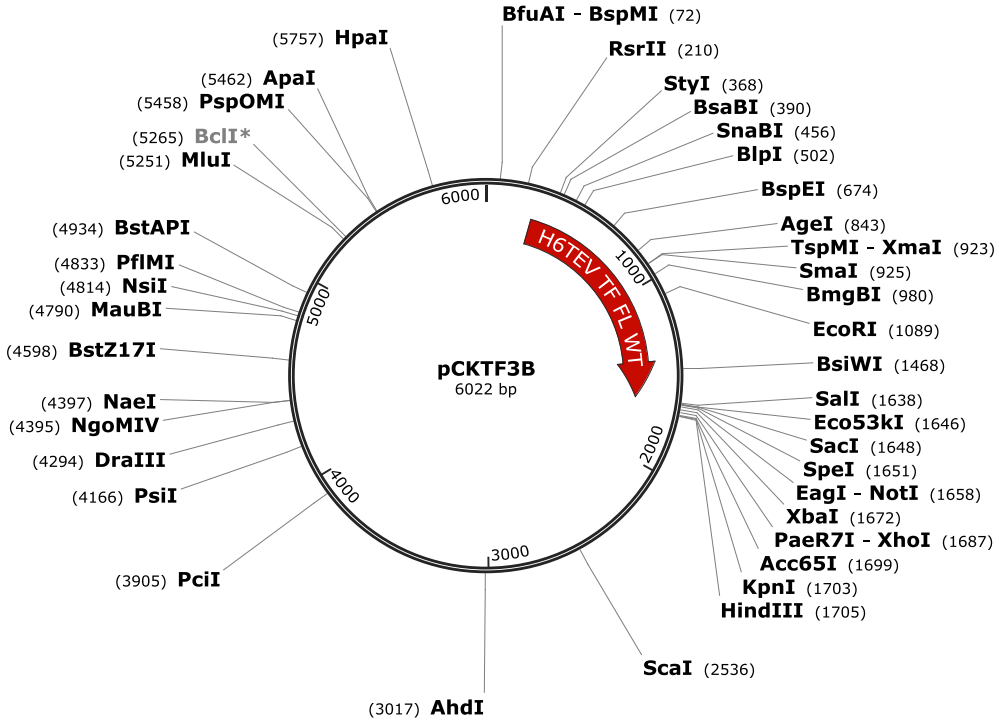
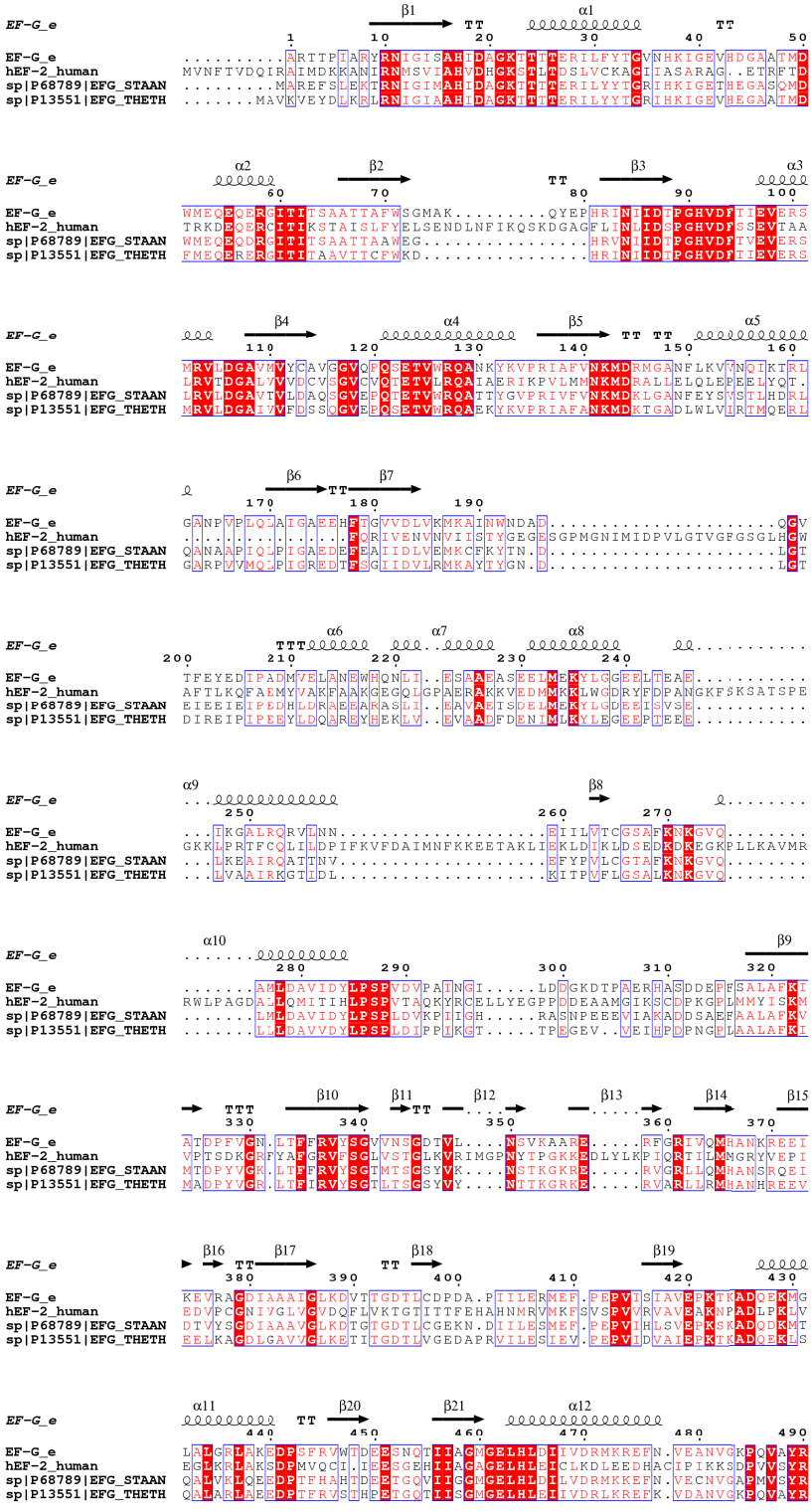


Figure-A7 Plasmid map of wild type trigger factor.

A-3 Multiple sequence alignment of EF-G



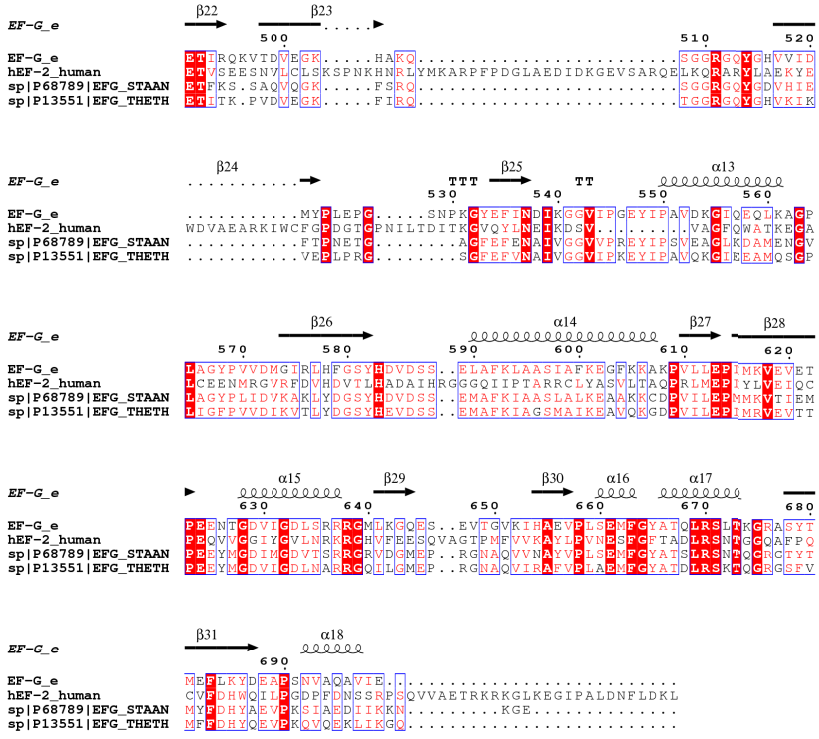


Figure-A8 Multiple-sequence alignment of EF-G. EF-G sequences are from different organisms: EF-G_e is from *Escherichia coli*, hEF-2_human is from *Homo sapiens*, sp|P68789|EFG_STAAN is from *staphylococcus* and sp|P13551|EFG_THETH is from *Thermus thermophilus*. The sequence alignment is done by Clustal Omega (Chojnacki et al., 2017). Figures were generated by ESPript 3.0 (Robert and Gouet, 2014).

References

AEvarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., Al-Karadaghi, S., Svensson, L.A., and Liljas, A. (1994). Three-dimensional structure of the ribosomal translocase: elongation factor G from Thermus thermophilus. EMBO J.

Agashe, V.R., Guha, S., Chang, H.C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U., and Barral, J.M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. Cell *117*, 199–209.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science (80-.). *181*, 223–230.

Batey, S., and Clarke, J. (2008). The Folding Pathway of a Single Domain in a Multidomain Protein is not Affected by Its Neighbouring Domain. J. Mol. Biol.

Batey, S., Randles, L.G., Steward, A., and Clarke, J. (2005). Cooperative folding in a multi-domain protein. J Mol Biol *349*, 1045–1059.

Batey, S., Scott, K.A., and Clarke, J. (2006). Complex folding kinetics of a multidomain protein. Biophys J *90*, 2120–2130.

Bauer, D., Merz, D.R., Pelz, B., Theisen, K.E., Yacyshyn, G., Mokranjac, D., Dima, R.I., Rief, M., Žoldák, G., and Zoldak, G. (2015). Nucleotides regulate the mechanical hierarchy between subdomains of the nucleotide binding domain of the Hsp70 chaperone DnaK. Proc. Natl. Acad. Sci. U. S. A. *112*, 10389.

Beckett, D., Kovaleva, E., and Schatz, P.J. (2008). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci.

Bell, G.I. (1978). Models for the specific adhesion of cells to cells. Science (80-.). *200*, 618–627.

Bhaskara, R.M., and Srinivasan, N. (2011). Stability of domain structures in multi-domain proteins. Sci Rep *1*, 40.

Borgia, A., Kemplen, K.R., Borgia, M.B., Soranno, A., Shammas, S., Wunderlich, B., Nettels, D., Best, R.B., Clarke, J., and Schuler, B. (2015). Transient misfolding dominates multidomain protein folding. Nat Commun *6*, 8861.

Borgia, M.B., Borgia, A., Best, R.B., Steward, A., Nettels, D., Wunderlich, B., Schuler, B., and Clarke, J. (2011). Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. Nature *474*, 662–665.

Boström, K., Wettesten, M., Borén, J., Bondjers, G., Wiklund, O., and Olofsson, S.O. (1986). Pulse-chase studies of the synthesis and intracellular transport of apolipoprotein B-100 in Hep G2 cells. J. Biol. Chem. *261*, 13800–13806.

Braselmann, E., Chaney, J.L., and Clark, P.L. (2013). Folding the proteome. Trends Biochem. Sci. *38*, 337–344.

Brockwell, D.J., and Radford, S.E. (2007). Intermediates: ubiquitous species on folding energy landscapes? Curr Opin Struct Biol *17*, 30–37.

Brockwell, D.J., Paci, E., Zinober, R.C., Beddard, G.S., Olmsted, P.D., Smith, D.A., Perham, R.N., and Radford, S.E. (2003). Pulling geometry defines the mechanical resistance of a β-sheet protein. Nat. Struct. Biol.

Bryngelson, J.D., and Wolynes, P.G. (1987). Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. *84*, 7524–7528.

Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M. V, and Komar, A.A. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein

Conformations. Mol Cell *61*, 341–351.

Bustamante, C., Marko, J.F., Siggia, E.D., and Smith, S. (1994). Entropic elasticity of lambda-phage DNA. Science (80-.). *265*, 1599–1600.

Bustamante, C., Chemla, Y.R., Forde, N.R., and Izhaky, D. (2004). Mechanical processes in biochemistry. Annu Rev Biochem *73*, 705–748.

Butkus, M.E., Prundeanu, L.B., and Oliver, D.B. (2003). Translocon "Pulling" of Nascent SecM Controls the Duration of Its Translational Pause and Secretion-Responsive secA Regulation. J. Bacteriol.

Cabrita, L.D., Danny Hsu, S.-T., Launay, H., Dobson, C.M., and Christodoulou, J. (2009). Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy.

Cabrita, L.D., Cassaignau, A.M., Launay, H.M., Waudby, C.A., Wlodarski, T., Camilloni, C., Karyadi, M.E., Robertson, A.L., Wang, X., Wentink, A.S., et al. (2016). A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. Nat Struct Mol Biol *23*, 278–285.

Calloni, G., Chen, T., Schermann, S.M., Chang, H., Genevaux, P., Agostini, F., Tartaglia, G.G., Hayer-Hartl, M., and Hartl, F.U. (2012). DnaK Functions as a Central Hub in the E. coli Chaperone Network. Cell Rep. *1*, 251–264.

Cecconi, C., Shank, E.A., Bustamante, C., and Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. Science (80-.). *309*, 2057–2060.

Chaney, J.L., Steele, A., Carmichael, R., Rodriguez, A., Specht, A.T., Ngo, K., Li, J., Emrich, S., and Clark, P.L. (2017). Widespread position-specific conservation of synonymous rare codons within coding sequences. PLoS Comput. Biol.

Chen, C.L., Cui, X.N., Beausang, J.F., Zhang, H.B., Farrell, I., Cooperman, B.S., and Goldman, Y.E. (2016). Elongation factor G initiates translocation through a power stroke. Proc. Natl. Acad. Sci. U. S. A. *113*, 7515–7520.

Chen, J., Sawyer, N., and Regan, L. (2013). Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. Protein Sci. *22*, 510–515.

Chodera, J., Elms, P., Noé, F., Keller, B., and John D. Chodera Frank Noé, Bettina Keller, Christian M. Kaiser, Aaron Ewall-Wice, Susan Marqusee, Carlos Bustamante, Nina Singhal Hinrichs, P.E. (2011). Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. ArXiv Prepr. ArXiv ….

Chojnacki, S., Cowley, A., Lee, J., Foix, A., and Lopez, R. (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. Nucleic Acids Res.

Clark, P.L., and King, J. (2001). A Newly Synthesized, Ribosome-bound Polypeptide Chain Adopts Conformations Dissimilar from Early in Vitro Refolding Intermediates. J. Biol. Chem.

Collin, D., Ritort, F., Jarzynski, C., Smith, S.B., Tinoco, I., and Bustamante, C. (2005). Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. Nature *437*, 231–234.

Crooks, G.E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top. *60*, 2721–2726.

Crowley, K.S., Liao, S., Worrell, V.E., Reinhart, G.D., and Johnson, A.E. (1994). Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore. Cell.

Cymer, F., and von Heijne, G. (2013). Cotranslational folding of membrane proteins probed by

arrest-peptide–mediated force measurements. Proc. Natl. Acad. Sci.

Czworkowski, J., Wang, J., Steitz, T.A., and Moore, P.B. (1994). The crystal structure of elongation factor G complexed with GDP, at 2.7 A resolution. EMBO J *13*, 3661–3668.

Dao Duc, K., and Song, Y.S. (2018). The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. PLoS Genet.

Deuerling, E., Patzelt, H., Vorderwulbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H., and Bukau, B. (2003). Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Mol Microbiol *47*, 1317–1328.

Dill, K.A., and MacCallum, J.L. (2012). The Protein-Folding Problem, 50 Years On. Science (80-.). *338*, 1042–1046.

Dobson, C.M. (2003). Protein folding and misfolding. Nature.

Dou, J., Vorobieva, A.A., Sheffler, W., Doyle, L.A., Park, H., Bick, M.J., Mao, B., Foight, G.W., Lee, M.Y., Gagnon, L.A., et al. (2018). De novo design of a fluorescence-activating β-barrel. Nature.

Dudko, O.K., Hummer, G., and Szabo, A. (2006). Intrinsic rates and activation free energies from single-molecule pulling experiments. Phys Rev Lett *96*, 108101.

Dudko, O.K., Hummer, G., and Szabo, A. (2008). Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. Proc Natl Acad Sci U S A *105*, 15755–15760.

Ellis, J.P., Bakke, C.K., Kirchdoerfer, R.N., Jungbauer, L.M., and Cavagnero, S. (2008). Chain dynamics of nascent polypeptides emerging from the ribosome. ACS Chem. Biol.

Ellis, J.P., Culviner, P.H., and Cavagnero, S. (2009). Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two

local motions. Protein Sci.

Evans, M.S., Sander, I.M., and Clark, P.L. (2008). Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. J Mol Biol *383*, 683–692.

Farías-Rico, J.A., Goetz, S.K., Marino, J., and von Heijne, G. (2017). Mutational analysis of protein folding inside the ribosome exit tunnel. FEBS Lett.

Fedorov, A.N., and Baldwin, T.O. (1995). Contribution of cotranslational folding to the rate of formation of native protein structure. Proc Natl Acad Sci U S A *92*, 1227–1231.

Fedorov, N.A., and Baldwin, O.T. (1997). Cotranslational Protein Folding - Minireview. J. Biol. Chem.

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature *431*, 590–596.

Flaugh, S.L. (2005). Contributions of hydrophobic domain interface interactions to the folding and stability of human D-crystallin. Protein Sci.

Frydman, J., Erdjument-Bromage, H., Tempst, P., and Ulrich Hartl, F. (1999). Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. Nat. Struct. Biol.

Genevaux, P., Keppel, F., Schwager, F., Langendijk-Genevaux, P.S., Hartl, F.U., and Georgopoulos, C. (2004). In vivo analysis of the overlapping functions of DnaK and trigger factor. EMBO Rep.

Goldman, D.H., Kaiser, C.M., Milin, A., Righini, M., Tinoco Jr., I., Bustamante, C., Tinoco, I., and Bustamante, C. (2015). Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. Science (80-.). *348*, 457–460.

Gupta, P., Liu, B., Klepacki, D., Gupta, V., Schulten, K., Mankin, A.S., and Vázquez-Laslop, N. (2016). Nascent peptide assists the ribosome in recognizing chemically distinct small molecules. Nat. Chem. Biol.

Haldar, S., Tapia-Rojo, R., Eckels, E.C., Valle-Orero, J., and Fernandez, J.M. (2017). Trigger factor chaperone acts as a mechanical foldase. Nat. Commun. *8*, 668.

Han, J.-H.H., Batey, S., Nickson, A.A., Teichmann, S.A., and Clarke, J. (2007). The folding and evolution of multidomain proteins. Nat Rev Mol Cell Biol *8*, 319–330.

Hartl, F.U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol *16*, 574–581.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324–332.

Heurgué-Hamard, V., Karimi, R., Mora, L., MacDougall, J., Leboeuf, C., Grentzmann, G., Ehrenberg, M., and Buckingham, R.H. (1998). Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome. EMBO J.

Himeno, H., Kurita, D., and Muto, A. (2014). TmRNA-mediated trans-translation as the major ribosome rescue system in a bacterial cell. Front. Genet.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. Biochim Biophys Acta *1803*, 650–661.

Hoffmann, A., Becker, A.H., Zachmann-Brand, B., Deuerling, E., Bukau, B., and Kramer, G. (2012). Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. Mol Cell *48*, 63–74.

Holtkamp, W., Kokic, G., Jager, M., Mittelstaet, J., Komar, A.A., Rodnina, M. V, Jäger, M.,

Mittelstaet, J., Komar, A.A., and Rodnina, M. V (2015). Cotranslational protein folding on the ribosome monitored in real time. Science (80-.). *350*, 1104–1107.

Horton, N., and Lewis, M. (1991). Calculation of the free energy of association for protein complexes. Protein Sci. *1*, 169–181.

Houwman, J.A., André, E., Westphal, A.H., Van Berkel, W.J.H., and Van Mierlo, C.P.M. (2016). The ribosome restrains molten globule formation in stalled nascent flavodoxin. J. Biol. Chem.

Hsu, S.T., Fucini, P., Cabrita, L.D., Launay, H., Dobson, C.M., and Christodoulou, J. (2007). Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. Proc Natl Acad Sci U S A *104*, 16516–16521.

Hummer, G., and Szabo, A. (2001). Free energy reconstruction from nonequilibrium single-molecule pulling experiments. Proc. Natl. Acad. Sci. U. S. A. *98*, 3658–3661.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Ishii, E., Chiba, S., Hashimoto, N., Kojima, S., Homma, M., Ito, K., Akiyama, Y., and Mori, H. (2015). Nascent chain-monitored remodeling of the Sec machinery for salinity adaptation of marine bacteria. Proc. Natl. Acad. Sci.

Ismail, N., Hedman, R., Schiller, N., and Von Heijne, G. (2012). A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. Nat. Struct. Mol. Biol.

Ito, K., and Chiba, S. (2013). Arrest Peptides: *Cis* -Acting Modulators of Translation. Annu. Rev. Biochem.

Jacobs, W.M., and Shakhnovich, E.I. (2017). Evidence of evolutionary selection for cotranslational folding. Proc Natl Acad Sci U S A *114*, 11434–11439.

Jahn, M., Buchner, J., Hugel, T., and Rief, M. (2016). Folding and assembly of the large molecular machine Hsp90 studied in single-molecule experiments. Proc Natl Acad Sci U S A *113*, 1232–1237.

Jahn, M., Tych, K., Girstmair, H., Steinmaßl, M., Hugel, T., Buchner, J., and Rief, M. (2018). Folding and Domain Interactions of Three Orthologs of Hsp90 Studied by Single-Molecule Force Spectroscopy. Structure *26*, 96-105.e4.

Jarzynski, C. (1997). Nonequilibrium Equality for Free Energy Differences. Phys. Rev. Lett. *78*, 2690–2693.

Kaiser, C.M., Chang, H.C., Agashe, V.R., Lakshmipathy, S.K., Etchells, S.A., Hayer-Hartl, M., Hartl, F.U., and Barral, J.M. (2006). Real-time observation of trigger factor function on translating ribosomes. Nature *444*, 455–460.

Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco Jr., I., and Bustamante, C. (2011). The ribosome modulates nascent protein folding. Science (80-.). *334*, 1723–1727.

Khushoo, A., Yang, Z., Johnson, A.E., and Skach, W.R. (2011). Ligand-driven vectorial folding of ribosome-bound human CFTR NBD1. Mol Cell *41*, 682–692.

Kim, S.J., Yoon, J.S., Shishido, H., Yang, Z., Rooney, L.A.A., Barral, J.M., and Skach, W.R. (2015). Translational tuning optimizes nascent protein folding in cells. Science (80-.). *348*.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013). Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem *82*, 323–355.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S. V, and Gottesman, M.M. (2007). A “silent” polymorphism in the MDR1 gene changes substrate specificity. Science (80-.). *315*, 525–528.

Knight, A.M., Culviner, P.H., Kurt-Yilmaz, N., Zou, T., Ozkan, S.B., and Cavagnero, S. (2013). Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. ACS Chem. Biol.

Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Palzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E., and Bukau, B. (2002). L23 protein functions as a chaperone docking site on the ribosome. Nature.

Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat Struct Mol Biol *16*, 589–597.

Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. J. Mol. Biol.

Kristensen, O., and Gajhede, M. (2003). Chaperone Binding at the Ribosomal Exit Tunnel. Structure.

Kubelka, J., Hofrichter, J., and Eaton, W.A. (2004). The protein folding "speed limit." Curr. Opin. Struct. Biol. *14*, 76–88.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. Science (80-.).

Kumar, V., and Chaudhuri, T.K. (2018). Spontaneous refolding of the large multidomain protein malate synthase G proceeds through misfolding traps. J. Biol. Chem.

Law, R., Carl, P., Harper, S., Dalhaimer, P., Speicher, D.W., and Discher, D.E. (2003). Cooperativity in forced unfolding of tandem spectrin repeats. Biophys J *84*, 533–544.

Levinthal, C. (1969). How to fold graciously. Mossbauer Spectrosc. ….

Lin, J., Gagnon, M.G., Bulkley, D., and Steitz, T.A. (2015). Conformational Changes of

Elongation Factor G on the Ribosome during tRNA Translocation. Cell *160*, 219–227.

Liphardt, J. (2007). Equilibrium Information from in an Experimental Test of Jarzynski ' s Equality. Science (80-.). *1832*, 1832–1835.

Liu, K., Rehfus, J.E., Mattson, E., and Kaiser, C.M. (2017). The ribosome destabilizes native and non-native structures in a nascent multidomain protein. Protein Sci. *26*, 1439–1451.

Liu, K., Maciuba, K., Kaiser, C.M., Liu, K., Maciuba, K., and Kaiser, C.M. (2019). The Ribosome Cooperates with a Chaperone to Guide Multi-domain Protein Folding Article The Ribosome Cooperates with a Chaperone to Guide Multi-domain Protein Folding. Mol. Cell *74*, 310–319.

Lu, J., and Deutsch, C. (2005). Folding zones inside the ribosomal exit tunnel. Nat. Struct. Mol. Biol.

Lu, P., Min, D., DiMaio, F., Wei, K.Y., Vahey, M.D., Boyken, S.E., Chen, Z., Fallas, J.A., Ueda, G., Sheffler, W., et al. (2018). Accurate computational design of multipass transmembrane proteins. Science (80-.).

Lucent, D., Snow, C.D., Aitken, C.E., and Pande, V.S. (2010). Non-bulk-like solvent behavior in the ribosome exit tunnel. PLoS Comput. Biol.

Marino, J., von Heijne, G., and Beckmann, R. (2016). Small protein domains fold inside the ribosome exit tunnel. FEBS Lett. *590*, 655–660.

Marsh, J.A., and Teichmann, S.A. (2014). Structure, Dynamics, Assembly, and Evolution of Protein Complexes. Annu. Rev. Biochem. *84*, 551–575.

Mashaghi, A., Kramer, G., Bechtluft, P., Zachmann-Brand, B., Driessen, A.J., Bukau, B., and Tans, S.J. (2013). Reshaping of the conformational search of a protein by the chaperone trigger factor. Nature *500*, 98–101.

Mashaghi, A., Kramer, G., Lamb, D.C., Mayer, M.P., and Tans, S.J. (2014). Chaperone action at the single-molecule level. Chem Rev *114*, 660–676.

Massey, F.J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. J. Am. Stat. Assoc.

Mercier, E., and Rodnina, M. V (2018). Co-Translational Folding Trajectory of the HemK Helical Domain. Biochemistry.

Mittal, J., and Best, R.B. (2008). Thermodynamics and kinetics of protein folding under confinement. Proc. Natl. Acad. Sci.

Mittelstaet, J., Konevega, A.L., and Rodnina, M. V. (2013). A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome. J. Am. Chem. Soc.

Mönkemeyer, L., Klaips, C.L., Balchin, D., Körner, R., Hartl, F.U., and Bracher, A. (2019). Chaperone Function of Hgh1 in the Biogenesis of Eukaryotic Elongation Factor 2. Mol. Cell 88–100.

Mossessova, E., and Lima, C.D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol. Cell.

Motlagh, H.N., Toptygin, D., Kaiser, C.M., and Hilser, V.J. (2016). Single-Molecule Chemo-Mechanical Spectroscopy Provides Structural Identity of Folding Intermediates. Biophys. J. *110*, 1280–1290.

Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. Cell.

Netzer, W.J., and Hartl, F.U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature *388*, 343–349.

Nicola, A. V., Chen, W., and Helenius, A. (1999). Co-translational folding of an alphavirus capsid

protein in the cytosol of living cells. Nat. Cell Biol. *1*, 341–345.

Nilsson, O.B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Müller-Lucks, A., Trovato, F., Puglisi, J.D., O'Brien, E.P., et al. (2015). Cotranslational Protein Folding inside the Ribosome Exit Tunnel. Cell Rep. *12*, 1533–1540.

Nilsson, O.B., Muller-Lucks, A., Kramer, G., Bukau, B., and von Heijne, G. (2016). Trigger Factor Reduces the Force Exerted on the Nascent Chain by a Cotranslationally Folding Protein. J Mol Biol *428*, 1356–1364.

Nilsson, O.B., Nickson, A.A., Hollins, J.J., Wickles, S., Steward, A., Beckmann, R., von Heijne, G., and Clarke, J. (2017). Cotranslational folding of spectrin domains via partially structured states. Nat. Struct. Mol. Biol. *24*, 221–225.

Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., et al. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell *147*, 1295–1308.

Onuchic, J.N., Luthey-Schulten, Z., and Wolynes, P.G. (1997). Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem *48*, 545–600.

Park, C., and Marqusee, S. (2005). Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. Nat. Methods.

Parker, M.J., Spencer, J., Jackson, G.S., Burston, S.G., Hosszu, L.L.P., Craven, C.J., Waltho, J.P., and Clarke, A.R. (1996). Domain behavior during the folding of a thermostable phosphoglycerate kinase. Biochemistry.

Peacock, J.A. (1983). Two-dimensional goodness-of-fit testing in astronomy. MNRAS.

Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals

hidden signatures of cotranslational folding. Nat. Struct. Mol. Biol. *20*, 237–243.

Phillips, D.C. (1967). THE HEN EGG-WHITE LYSOZYME MOLECULE. Proc. Natl. Acad. Sci. *57*, 483–495.

Preissler, S., and Deuerling, E. (2012). Ribosome-associated chaperones as key players in proteostasis. Trends Biochem. Sci.

Pulk, A., and Cate, J.H. (2013). Control of ribosomal subunit rotation by elongation factor G. Science (80-.). *340*, 1235970.

Randles, L.G., Batey, S., Steward, A., and Clarke, J. (2008). Distinguishing specific and nonspecific interdomain interactions in multidomain proteins. Biophys J *94*, 622–628.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res.

Rodnina, M. V. (2016). The ribosome in action: Tuning of translational efficiency and protein folding. Protein Sci. *25*, 1390–1406.

Rodnina, M. V, and Wintermeyer, W. (2016). Protein Elongation, Co-translational Folding and Targeting. J Mol Biol *428*, 2165–2185.

Rudolph, R., Siebendritt, R., Nesslauer, G., Sharma, A.K., and Jaenicke, R. (1990). Folding of an all-β protein: Independent domain in γII-crystallin from calf eye lens. Proc Natl Acad Sci USA *87*, 4625–4629.

Saio, T., Guan, X., Rossi, P., Economou, A., and Kalodimos, C.G. (2014). Structural basis for protein antiaggregation activity of the trigger factor chaperone. Science (80-.). *344*, 1250494.

Samelson, A.J., Jensen, M.K., Soto, R.A., Cate, J.H.D., and Marqusee, S. (2016). Quantitative determination of ribosome nascent chain stability. Proc. Natl. Acad. Sci. *113*, 13402–13407.

Sánchez, I.E., Morillas, M., Zobeley, E., Kiefhaber, T., and Glockshuber, R. (2004). Fast folding of the two-domain Semliki Forest virus capsid protein explains co-translational proteolytic activity. J. Mol. Biol. *338*, 159–167.

Sander, I.M., Chaney, J.L., and Clark, P.L. (2014). Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. J Am Chem Soc *136*, 858–861.

Schaffitzel, C., and Ban, N. (2007). Generation of ribosome nascent chain complexes for structural and functional studies. J. Struct. Biol.

Scholl, Z.N., Yang, W., and Marszalek, P.E. (2014). Chaperones rescue luciferase folding by separating its domains. J Biol Chem *289*, 28607–28618.

Scholl, Z.N., Yang, W., and Marszalek, P.E. (2017). Competing Pathways and Multiple Folding Nuclei in a Large Multidomain Protein, Luciferase. Biophys J *112*, 1829–1840.

Schuler, B., and Hofmann, H. (2013). Single-molecule spectroscopy of protein folding dynamics--expanding scope and timescales. Curr Opin Struct Biol *23*, 36–47.

Scott, K.A., Steward, A., Fowler, S.B., and Clarke, J. (2002). Titin; a multidomain protein that behaves as the sum of its parts. J. Mol. Biol. *315*, 819–829.

Sgouralis, I., Whitmore, M., Lapidus, L., Comstock, M.J., and Pressé, S. (2018). Single molecule force spectroscopy at high data acquisition: A Bayesian nonparametric analysis. J. Chem. Phys. *148*, 1–13.

Shank, E.A., Cecconi, C., Dill, J.W., Marqusee, S., and Bustamante, C. (2010). The folding cooperativity of a protein is controlled by its chain topology. Nat. 2010 4657298 *465*, 637.

Shieh, Y.-W., Minguez, P., Bork, P., Auburger, J.J., Guilbride, D.L., Kramer, G., and Bukau, B. (2015). Operon structure and cotranslational subunit association direct protein assembly in bacteria.

Science (80-.). *350*, 678–680.

Siller, E., DeZwaan, D.C., Anderson, J.F., Freeman, B.C., and Barral, J.M. (2010). Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. J Mol Biol *396*, 1310–1318.

Smith, S.B., Cui, Y., and Bustamante, C. (1996). Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. Science (80-.).

Smith, S.B., Cui, Y.J., and Bustamante, C. (2003). Optical-trap force transducer that operates by direct measurement of light momentum. Biophotonics, Pt B *361*, 134–162.

Sosnick, T.R., and Barrick, D. (2011). The folding of single domain proteins--have we reached a consensus? Curr Opin Struct Biol *21*, 12–24.

Spencer, P.S., Siller, E., Anderson, J.F., and Barral, J.M. (2012). Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. J Mol Biol *422*, 328–335.

Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J.C., and Rief, M. (2011). The complex folding network of single calmodulin molecules. Science (80-.). *334*, 512–516.

Su, T., Cheng, J., Sohmen, D., Hedman, R., Berninghausen, O., von Heijne, G., Wilson, D.N., and Beckmann, R. (2017). The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. Elife.

Thommen, M., Holtkamp, W., and Rodnina, M. V. (2017). Co-translational protein folding: progress and methods. Curr. Opin. Struct. Biol.

Tugarinov, V., and Kay, L.E. (2003). Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. J. Am. Chem. Soc.

Tzul, F.O., Vasilchuk, D., and Makhatadze, G.I. (2017). Evidence for the principle of minimal

frustration in the evolution of protein folding landscapes. Proc Natl Acad Sci U S A.

Ugrinov, K.G., and Clark, P.L. (2010). Cotranslational folding increases GFP folding yield. Biophys J *98*, 1312–1320.

Vorderwulbecke, S., Kramer, G., Merz, F., Kurz, T.A., Rauch, T., Zachmann-Brand, B., Bukau, B., and Deuerling, E. (2004). Low temperature or GroEL/ES overproduction permits growth of Escherichia coli cells lacking trigger factor and DnaK. FEBS Lett *559*, 181–187.

Walters, B.T., Mayne, L., Hinshaw, J.R., Sosnick, T.R., and Englander, S.W. (2013). Folding of a large protein at high structural resolution. Proc. Natl. Acad. Sci. U. S. A. *110*, 18898–18903.

Wilson, D.N., Arenz, S., and Beckmann, R. (2016). Translation regulation via nascent polypeptide-mediated ribosome stalling The nascent polypeptide chain within the ribosomal tunnel. Curr. Opin. Struct. Biol. *37*, 123–133.

Woodside, M.T., and Block, S.M. (2014). Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu Rev Biophys *43*, 19–39.

Woolhead, C.A., McCormick, P.J., and Johnson, A.E. (2004). Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. Cell.

Woolhead, C.A., Johnson, A.E., and Bernstein, H.D. (2006). Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. Mol Cell *22*, 587–598.

Yamamoto, H., Qin, Y., Achenbach, J., Li, C., Kijek, J., Spahn, C.M., and Nierhaus, K.H. (2014). EF-G and EF4: translocation and back-translocation on the bacterial ribosome. Nat Rev Microbiol *12*, 89–100.

Yin, J., Lin, A.J., Golan, D.E., and Walsh, C.T. (2006). Site-specific protein labeling by Sfp

phosphopantetheinyl transferase. Nat. Protoc.

Young, R., and Bremer, H. (1976). Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. Biochem. J. *160*, 185–194.

Young, T.A., Skordalakes, E., and Marqusee, S. (2007). Comparison of Proteolytic Susceptibility in Phosphoglycerate Kinases from Yeast and E. coli: Modulation of Conformational Ensembles Without Altering Structure or Stability. J. Mol. Biol. *368*, 1438–1447.

Yu, H., Liu, X., Neupane, K., Gupta, A.N., Brigley, A.M., Solanki, A., Sosova, I., and Woodside, M.T. (2012). Direct observation of multiple misfolding pathways in a single prion protein molecule. Proc Natl Acad Sci U S A *109*, 5283–5288.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol *16*, 274–280.

Zheng, W., Schafer, N.P., and Wolynes, P.G. (2013). Frustration in the energy landscapes of multidomain protein misfolding. Proc Natl Acad Sci U S A *110*, 1680–1685.

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J.M., Sachs, M.S., and Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature.

Zhuravlev, P.I., Hinczewski, M., Chakrabarti, S., Marqusee, S., and Thirumalai, D. (2016). Force-dependent switch in protein unfolding pathways and transition-state movements. Proc Natl Acad Sci U S A *113*, E715-24.

Zoldak, G., and Rief, M. (2013). Force as a single molecule probe of multidimensional protein energy landscapes. Curr. Opin. Struct. Biol. *23*, 48–57.

Curriculum Vitae

Kaixian Liu

Ph.D. candidate

Contact information

Department of Biology
Johns Hopkins University
3400 North Charles Street, 209 Mudd Hall
Baltimore, MD 21218-2685

Phone (cell): 443-240-7960
email: kliu39@jhu.edu
kaixianliu@gmail.com

Brief Bio

Kaixian Liu was born at Wannian, Jiangxi province, China on May 20th, 1989.

Education

2019, expected *Ph.D.*, Cellular, Molecular and Developmental Biology and Biophysics, Johns Hopkins University

2012 *M.S.* of Structural Biology
University of Science and Technology of China

2010 *B.S.* of Biology
University of Science and Technology of China

Research Experience

2013-Present Doctoral Thesis Research

Department of Biology

Johns Hopkins University

Advisor: Christian Kaiser

Title: *Folding of a multidomain protein, with the help of the ribosome and a molecular chaperone*

2010-2012 Graduate Research

School of life sciences

University of Science and Technology of China

Advisors: Yunyu Shi and Jihui Wu

Title: *NMR studies on the interface of human Mog1 and Ran*

2009-2010 Undergraduate Research

School of life sciences

University of Science and Technology of China

Advisors: Yunyu Shi and Jihui Wu

Title: *Structural studies on tandem BRCT repeats of Rtt107*

Teaching and Supervisory Experience

2014-2015 Teaching assistant for undergraduate laboratory classes

Cell Biology and Biochemistry

2017 Teaching assistant for undergraduate and graduate class

Single-Molecule Approaches to Biology

2018 Teaching assistant for undergraduate class

Comparative Physiology

2014-2018 Supervise 7 graduate students and 1 undergraduate student

Oral Presentations

1. Inter-domain interactions in nascent polypeptides interfere with productive folding. (2017)

The 1st Chesapeake Bay Area Single Molecule Biology Meeting, Baltimore, MD**.**

2. Inter-domain interactions in nascent polypeptides interfere with productive folding. (2017)

2017 Annual meeting of American Society for Biochemistry and Molecular Biology, Chicago, IL**.**

Poster Presentations

1. Tuned energetics of domain-domain interactions in a complex protein. (2019)

The 10th Biennial Single Molecule Biophysics Meeting, Aspen, CO.

2. The ribosome and the chaperone trigger factor guide multi-domain nascent protein folding co-translationally. (2018)

The 3rd Chesapeake Bay Area Single Molecule Biology Meeting, Baltimore, MD**.**

3. Inter-domain interactions in nascent polypeptides interfere with productive folding. (2017)

2017 Annual Meeting of American Society for Biochemistry and Molecular Biology, Chicago, IL**.**

4. Single-molecule studies of multi-domain protein folding. (2015)

The 59th Annual Meeting of the Biophysical Society, Baltimore, MD**.**

Publications

1. **Liu, K.**, Maciuba, K. and Kaiser, C. M. (2019). The ribosome cooperates with a chaperone to guide multi-domain protein folding. *Molecular Cell*, 74, 310-319.
2. Kaiser, C.M., and **Liu, K.** (2018). Folding up and moving on-nascent protein folding on the ribosome. *J. Mol. Biol.*, 430 (22):4580-4591.
3. **Liu, K.**, Rehfus, J. E., Mattson, E. and Kaiser, C. M. (2017). The ribosome destabilizes native and non-native structures in a nascent multidomain protein. *Protein Science*, *26*, 1439–1451.
4. Li, J., White, J. T., Saavedra, H., Wrabl, J. O., Motlagh, H. N., **Liu, K.**, … Hilser, V. J. (2017). Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. *eLife*, *6*, e30688.
5. Li, X, **Liu, K.**, Li, F, et al (2012). Structure of C-terminal Tandem BRCT Repeats of Rtt107 Protein Reveals Critical Role in Interaction with Phosphorylated Histone H2A during DNA Damage Repair. *The Journal of Biological Chemistry*, 287 (12):9137-9146.